The American
Economic Review

391
.001

78

Cuk-H02195-78-pd17822

ARTICLES

R. L. JACOBS AND R. A. JONES
Price Expectations in the United States, 1947–75

R. N. BATRA AND R. RAMACHANDRAN
Multinational Firms and the Theory of International Trade and Investment

R. FINDLAY
The Terms of Trade and Equilibrium Growth in the World Economy

J. E. HARRIS
Taxing Tar and Nicotine

A. DEATON AND J. MUELLBAUER
An Almost Ideal Demand System

S. A. HOENACK AND W. C. WEILER
A Structural Model of Murder Behavior

H. S. BURNESS, W. D. MONTGOMERY, AND J. P. QUIRK
Capital Contracting and the Regulated Firm

F. P. STAFFORD AND G. J. DUNCAN
Do Union Members Receive Compensating Wage Differentials?

M. J. BAILEY, M. OLSON, AND P. WONNACOTT
Marginal Utility of Income Does Not Increase: Borrowing, Lending, and Friedman-Savage Gambles

D. A. STARRETT
Taxation and the Provision of Local Public Goods

S. J. GROSSMAN AND J. E. STIGLITZ
The Impossibility of Informationally Efficient Markets

R. T. FROYEN AND R. N. WAUD
International Evidence on Output-Inflation Tradeoffs

L. R. CHRISTENSEN AND D. W. CAVES
Global Properties of Flexible Functional Forms

SHORTER PAPERS: M. Farrell; S. Oster; A. Randall and J. R. Stoll; S. P. Das; R. J. Staaf and E. G. West; A. T. Denzau and R. J. Mackay; R. Halvorsen and R. Palmquist; N. L. Jacob and A. N. Page; J. P. Formby and T. G. Seaks; R. Bradburd; G. Philpotts; J. J. Merrick, Jr.; J. W. Elliott; T. S. McCaleb; E. Helpman and D. Pines; H. Akiba; J. P. Hughes; L. Young; S. P. Das.

JUNE 1980

THE AMERICAN ECONOMIC ASSOCIATION

- Published at George Banta Co., Inc. Menasha, Wisconsin. The publication number is ISSN 0002-8282.

- *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Effective April 1, 1980, Association membership dues for 1980, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$28.75 for regular members with rank of assistant professor or lower, or with annual income of $13,800 or less;

$34.50 for regular members with rank of associate professor, or with annual income of $13,800 to $23,000;

$40.25 for regular members with rank of full professor, or with annual income above $23,000;

$14.25 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $43.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $4.75 to cover extra postage.

- Correspondence relating to the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

- Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212.

Founded in 1885

Officers

President

MOSES ABRAMOVITZ
Stanford University

President-Elect

WILLIAM J. BAUMOL
Princeton University and New York University

Vice Presidents

CARL F. CHRIST
Johns Hopkins University
H. GREGG LEWIS
Duke University

Secretary

C. ELTON HINSHAW
Vanderbilt University

Treasurer

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

ROBERT W. CLOWER
University of California-Los Angeles
MARINA v. N. WHITMAN
General Motors Corporation
HENRY J. AARON
The Brookings Institution
ZVI GRILICHES
Harvard University
MARTIN S. FELDSTEIN
National Bureau of Economic Research, Inc. and Harvard University
ROBERT E. LUCAS, JR.
University of Chicago

Ex Officio Members

TJALLING KOOPMANS
Yale University
ROBERT M. SOLOW
Massachusetts Institute of Technology

70

RONALD COASE

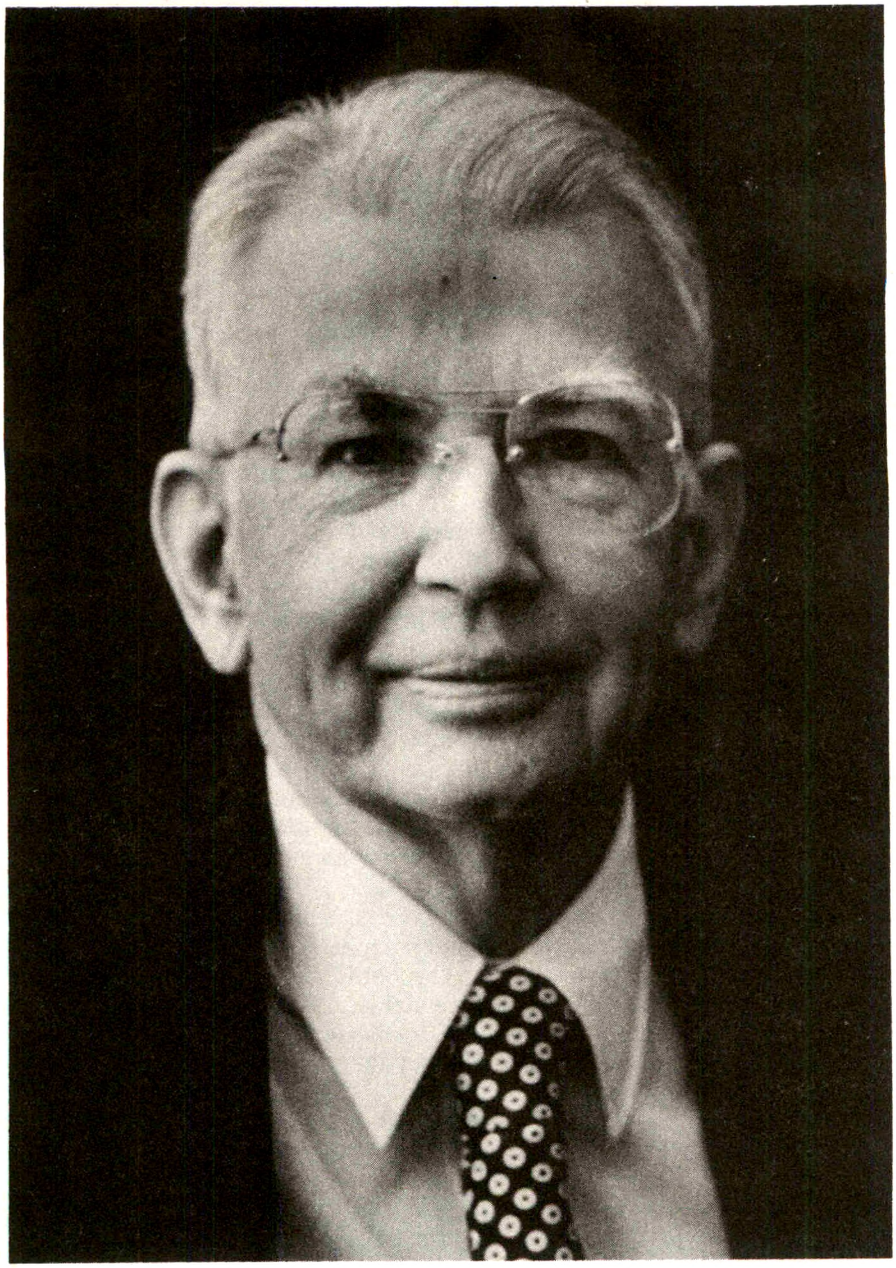
DISTINGUISHED FELLOW
1980

In a generation that will be remembered for its introduction of logical rigor into economics and for its refinement of quantitative empirical tools, a few economists have steadfastly practiced economics as a social science, a study of relationships among men mediated by the social institutions through which they deal with one another.

Ronald Coase stands in the forefront of these economists. He marked out his concern with the influence of institutions on economic behavior early in this career, raising questions about phenomena that his predecessors had taken for granted, such as what determines which transactions are conducted within firms and which by market purchase and sale. Not long after, in writing on the paradoxes that arise from marginal analysis where there are increasing returns to scale, Coase went beyond local considerations and analyzed the global aspects and the institutional complications entailed by charging prices that do not cover total costs.

From those days to this, Coase has persisted in his acute analysis of how institutions work in the economic sphere. He showed that externalities could not exist without the transactions costs imposed by the institutional arrangements inevitable in real economies. He pioneered in exploring the economic implications of the law and legal institutions. These papers do not fade; they grow in pertinence and influence over the years.

Antaeus lost his strength when his feet left the ground. We honor Ronald Coase for keeping his feet always grounded in the realities of economic-social life, and for reminding us, by demonstrating the strength so gained, that that is where our feet belong also.



Ronald H Coase

THE AMERICAN ECONOMIC REVIEW

June 1980

VOLUME 70, NUMBER 3

391
.001

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors
ALBERT ANDO
PRANAB K. BARDHAN
PETER A. DIAMOND
W. ERWIN DIEWERT
RUDIGER DORNBUSCH
HERSCHEL I. GROSSMAN
PETER W. HOWITT
ANNE O. KRUEGER
WILLIAM H. OAKLAND
MICHAEL PARKIN
ROY RADNER
RICHARD W. ROLL
NANCY L. SCHWARTZ
JAMES P. SMITH
A. MICHAEL SPENCE
WILLIAM S. VICKREY
ROBERT D. WILLIG
S. Y. WU

●Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to George H. Borts, Managing Editor, Box Q, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

●No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

●Copyright American Economic Association 1980. All rights reserved.

Articles

Price Expectations in the United States: 1947–75
Rodney L. Jacobs and Robert A. Jones 269

Multinational Firms and the Theory of International Trade and Investment
Raveendra N. Batra and Rama Ramachandran 278

The Terms of Trade and Equilibrium Growth in the World Economy *Ronald Findlay* 291

Taxing Tar and Nicotine *Jeffrey E. Harris* 300

An Almost Ideal Demand System
Angus Deaton and John Muellbauer 312

A Structural Model of Murder Behavior and the Criminal Justice System
Stephen A. Hoenack and William C. Weiler 327

Capital Contracting and the Regulated Firm
H. Stuart Burness, W. David Montgomery, and James P. Quirk 342

Do Union Members Receive Compensating Wage Differentials?
Greg. J. Duncan and Frank P. Stafford 355

The Marginal Utility of Income Does Not Increase: Borrowing, Lending, and Friedman-Savage Gambles
Martin J. Bailey, Mancur Olson, and Paul Wonnacott 372

On the Method of Taxation and the Provision of Local Public Goods *David A. Starrett* 380

On the Impossibility of Informationally Efficient Markets
Sanford J. Grossman and Joseph E. Stiglitz 393

Further International Evidence on Output-Inflation Tradeoffs
Richard T. Froyen and Roger N. Waud 409

Global Properties of Flexible Functional Forms
Douglas W. Caves and Laurits R. Christensen 422

Shorter Papers

International Impact of U.S. Money Supply: The Case of Mexico *Mike Farrell* 433
The Optimal Order for Submitting Manuscripts *Sharon Oster* 444
Consumer's Surplus in Commodity Space *Alan Randall and John R. Stoll* 449
On the Effect of Rate-of-Return Regulation under Uncertainty *Satya P. Das* 456
Limits on Public Provisions of Private Goods:
(Comment) *Robert J. Staaf and E. G. West* 461
Reply and Further Results *Arthur T. Denzau and Robert J. Mackay* 466
The Interpretation of Dummy Variables in Semilogarithmic Equations *Robert Halvorsen and Raymond Palmquist* 474
Production, Information Costs, and Economic Organization: The Buyer Monitoring Case *Nancy L. Jacob and Alfred N. Page* 476
Paglin's Gini Measure of Inequality: A Modification *John P. Formby and Terry G. Seaks* 479
Conglomerate Power without Market Power: The Effects of Conglomeration on a Risk-Averse Quantity-Adjusting Firm *Ralph M. Bradburd* 483
A Note on the Representation of Preferences in the Lindahl-Johansen Diagram *Geoffrey Philpotts* 488
Measuring the Expected Real Rate of Interest—An Exploration of Macro-Economic Alternatives:
Comment *John J. Merrick, Jr.* 493
Reply and Further Thoughts *J. Walter Elliott* 498
Excess Burden, Benefit Taxation, and Efficiency in Public Expenditure *Thomas S. McCaleb* 501
Optimal Public Investment and Dispersion Policy in a System of Open Cities *Elhanan Helpman and David Pines* 507
Gains from Trade under Uncertainty: Further Comment *Hiroya Akiba* 515
The Comparative Statics of the Competitive, Increasing-Cost Industry *Joseph P. Hughes* 518
Tarriffs vs. Quotas under Uncertainty: An Extension *Leslie Young* 522
Further Results on Input Choices under Uncertain Demand *Satya P. Das* 528

Notes 533

Price Expectations in the United States: 1947–75

By RODNEY L. JACOBS AND ROBERT A. JONES*

It is common to assume that expectations about future inflation influence economic behavior. The behavior of an economic system over time is determined in part by the manner in which these expectations are formed. Theoretical and empirical investigations frequently utilize a hypothesis that agents forecast future inflation rates primarily on the basis of past inflation rates: static, adaptive, extrapolative, and regressive expectations adjustment schemes are examples of such forecasting rules.[1] But such rules generally ignore the fact that individuals observe prices, not inflation rates. A change in observed prices can result from general price inflation, but it may also result from transitory shocks to the price level or observation error. Similarly, apparent changes in the general inflation rate may be purely transitory in nature or may signify the beginning of a trend. A model of inflation expectations should admit the possibility of these different sources of price change and embody the natural human tendency to extrapolate seeming trends.

This paper examines the extent to which agents' reported price expectations are consistent with such "naive" forecasting. A multilevel adaptive expectations model is developed that takes observed prices as the information available to agents. On the basis of these prices, individuals revise their beliefs about not only the price level but also the underlying inflation rate and trend in the inflation rate. When fitted to the Livingston survey data the model appears to provide a unified explanation of price expectations in the United States from 1947 to 1975. This is contrasted with earlier investigations suggesting that expectations formation differed significantly in the periods before and after 1960.

I. The Adaptive Learning Model

This section develops an algorithm by which agents might update their expectations about the price level, inflation rate, and higher order rates of change of prices. Consider first the standard adaptive expectations learning rule

$$(1) \qquad P^e_{t+1} = P^e_t + \lambda_1 (P_t - P^e_t)$$

where P_t is the observed price at time t, P^e_t is the forecast of P_t made at time $t-1$, λ_1 is a constant positive adaptation coefficient, and P^e_{t+1} is the forecast of P_{t+1} made after observing P_t. John Muth demonstrates that (1) provides a minimum mean squared error forecast, relative to the class of all linear constant coefficient functions of past observations of P, if P_t is generated by the stochastic process

$$(2) \qquad \begin{aligned} P_t &= p_t + u_t \\ p_t &= p_{t-1} + v_t \end{aligned}$$

where u_t and v_t are independent random processes with means 0 and respective variances μ and ν; p_t is interpreted as the true underlying price level; u_t as a transitory shock to prices (or observation error); and v_t as a permanent shock to the price level.

---

*Assistant professors of economics, University of California-Los Angeles. We wish to thank John Carlson for helpful comments on an earlier version, and Jon Salmon for research assistance.

[1] Alternatively, one might hypothesize that agents form their forecasts rationally on the basis of all economic variables known to them and with full knowledge of the economic process generating inflation. This hypothesis has considerable theoretical appeal as a means of consistently closing a model with endogenous expectations, but has been criticized for ignoring the substantial costs to an individual of acquiring and processing large quantities of data. Naive forecasting rules based only on past observations may be "economically rational" (see Edgar Feige and Douglas Pearce) if the value of improved prediction from more sophisticated rules is outweighed by the cost of collecting and processing the additional information. Moreover, the rational and the seemingly naive forecasting rules may be equivalent in some circumstances (see Michael Mussa).

Persistent trends in the price level are not likely to be generated by the process (2). Yet upward trends in prices clearly prevail during periods of inflation. In the presence of trends in P_t, the simple adaptive rule (1) results in consistent under- or overprediction of prices, and would lead a rational individual to reject the working hypothesis that process (2) generates prices in favor of a process in which trends are more likely.

A more general process, which does admit the possibility of trends which change over time, is

$$
\begin{aligned}
(3)\qquad P_t &= p_t + u_t \\
p_t &= p_{t-1} + \pi_t + v_t \\
\pi_t &= \pi_{t-1} + w_t
\end{aligned}
$$

where u_t, v_t, and w_t are independent white noise processes with means 0 and respective variances μ, ν, and ω. The variable P_t is still the observed price and p_t the true underlying price, but π_t is the true underlying inflation rate with v_t being a transitory shock and w_t a permanent shock to the trend in prices.[2] With π_0 and ω both equal to zero, process (3) reduces to process (2). With ν equal to zero, process (3) becomes that used by Marc Nerlove and S. Wage to demonstrate the optimality of adaptive forecasting. With μ equal to zero, and $\Pi_t \equiv P_t - P_{t-1}$ defining the observed inflation rate, process (3) becomes process (2) with observed inflation Π_t replacing observed prices P_t and with the underlying inflation rate π_t replacing underlying prices p_t.

A minimum mean squared error forecast for P_{t+1} when process (3) generates prices is provided by the two-level adaptive rule[3]

$$
\begin{aligned}
(4)\qquad P^e_{t+1} &= P^e_t + \lambda_1(P_t - P^e_t) + \Pi^e_{t+1} \\
\Pi^e_{t+1} &= \Pi^e_t + \lambda_2(P_t - P^e_t)
\end{aligned}
$$

The adaptation coefficients are positive constants whose values depend on the relative magnitudes of μ, ν, and ω. The variable Π^e_{t+1} is the expected *underlying* inflation rate over the period from t to t+1. This will differ from the expected *observed* inflation rate $P^e_{t+1} - P_t$ over the same period by the estimated current transitory shock to prices $P_t - [P^e_t + \lambda_1(P_t - P^e_t)]$. Of course if $\mu = 0$ then $\lambda_1 = 1$, and (4) reduces to the usual adaptive rule on observed inflation rate[4]

$$
(5)\qquad \Pi^e_{t+1} = \Pi^e_t + \lambda_2(\Pi_t - \Pi^e_t)
$$

Learning rule (4) solves the problem of consistently biased price forecasts in the presence of underlying inflation by having individuals update a forecast of both the price level and its trend. Suppose, for example, that the observed inflation rate jumps from 0 to higher constant value π. The individual would initially attribute much of the prediction error to transitory shocks in the price level and underlying inflation rate. This would lead to forecasts P^e_{t+1} which turn out consistently below P_{t+1}, and induce him to revise upward his estimate Π^e of the trend in prices. In general, Π^e overshoots the true trend, either converging to π through a damped oscillation or being critically damped to π.[5]

Periods of persistently rising or falling inflation rates have also been known to occur. How would prediction rule (4) fare in such situations? During the periods of persistent drift in Π, rule (4) consistently under- or overpredicts the price level, and would lead a rational individual to reject the hypothesis that (3) represents the process generating prices. A still more general process, which admits the possibility of a persistent drift d_t in the underlying inflation rate, is

---

[2] π_t is actually the rate of change in the price level p_t. Only if p_t denotes the logarithm of the price level, as it will for estimation purposes later in the paper, does π_t represent the inflation rate in the sense of the compound growth rate of prices.

[3] In our 1977 paper, we derive these equations as the asymptotic form of a Bayesian revision rule.

[4] The first equation of (4) lagged one period becomes $P^e_t = P_{t-1} + \Pi^e_t$, which when substituted into the equation for Π^e_{t+1} gives (5). The first equation of (4) is no longer needed for the expectation algorithm.

[5] The path of Π^e depends on the roots of the characteristic equation $\rho^2 - \rho(2-\lambda_1-\lambda_2) + (1-\lambda_1) = 0$. The model is stable if the roots lie within the unit circle.

$$(6)\qquad \begin{aligned} P_t &= p_t + u_t \\ p_t &= p_{t-1} + \pi_t + v_t \\ \pi_t &= \pi_{t-1} + d_t + w_t \\ d_t &= d_{t-1} + z_t \end{aligned}$$

in which u_t, v_t, w_t, and z_t are independent white noise processes. The minimum mean squared error forecast of P_{t+1} for process (6) is generated by the three-level adaptive learning rule

$$(7)\qquad \begin{aligned} P^e_{t+1} &= P^e_t + \lambda_1(P_t - P^e_t) + \Pi^e_{t+1} \\ \Pi^e_{t+1} &= \Pi^e_t + \lambda_2(P_t - P^e_t) + d^e_{t+1} \\ d^e_{t+1} &= d^e_t + \lambda_3(P_t - P^e_t) \end{aligned}$$

If there are no transitory shocks to the price level ($\mu = 0$) then $\lambda_1 = 1$. The distinction between whether prices P_t or inflation rates Π_t are the primary observation then vanishes since (7) can be rewritten as

$$(8)\qquad \begin{aligned} \Pi^e_{t+1} &= \Pi^e_t + \lambda_2(\Pi_t - \Pi^e_t) + d^e_{t+1} \\ d^e_{t+1} &= d^e_t + \lambda_3(\Pi_t - \Pi^e_t) \end{aligned}$$

The three-level adaptive learning model (7) is fitted to survey data on expected prices in the next section. Some of its properties warrant note at this point. The reader may verify that the prediction errors asymptotically approach zero if either the price level, the inflation rate, or the rate of change in the inflation rate is constant over time.[6] Thus there is an implicit tendency in the model to eventually extrapolate changes in inflation when $\lambda_3 > 0$. Jacob Frenkel, however, suggests that inflation expectations might be regressive over short periods of time, which he uses to explain the rise in real money balances observed in the early phase of monetary expansions. Is (7) consistent with short-run regressivity of inflation expectations as well as long-run extrapolativity? Suppose the inflation rate has been zero and that P^e_1, Π^e_1, and d^e_1 are fully adjusted to initial values of P_0, 0, and 0, respectively. Then a higher price level P_1 (equivalently, positive inflation of $\Pi_1 = P_1 - P_0$) is observed. The price level predicted by (7) for the following period is $P^e_2 = P_0 + (\lambda_1 + \lambda_2 + \lambda_3)(P_1 - P_0)$. The expected *observed* inflation rate $(P^e_2 - P_1)$ is hence $(\lambda_1 + \lambda_2 + \lambda_3 - 1)(P_1 - P_0)$ or $(\lambda_1 + \lambda_2 + \lambda_3 - 1)\Pi_1$. Although higher-than-expected prices always cause an upward revision in the believed *underlying* inflation rate, expectations can be regressive in the sense that the expected *observed* inflation rate initially moves in the opposite direction if $\lambda_1 + \lambda_2 + \lambda_3 < 1$. Short-run regressive and long-run extrapolative forecasting of observed inflation are neither mutually exclusive in nor a priori implied by (7). They correspond to the testable hypotheses that $\lambda_1 + \lambda_2 + \lambda_3 < 1$ and $\lambda_3 > 0$, respectively.[7]

Estimation of the multilevel adaptive expectations model requires the solving of difference equation system (7). Defining the column vectors

$$X_t = \begin{bmatrix} P^e \\ \Pi^e \\ d^e \end{bmatrix}_{t+1} \quad \text{and} \quad Z_t = \begin{bmatrix} \lambda_1 + \lambda_2 + \lambda_3 \\ \lambda_2 + \lambda_3 \\ \lambda_3 \end{bmatrix} P_t$$

the model can be expressed in matrix form:

$$(9)\qquad X_t = AX_{t-1} + Z_t \quad \text{with}$$

$$A = \begin{bmatrix} 1-\lambda_1-\lambda_2-\lambda_3 & 1.0 & 1.0 \\ -\lambda_2-\lambda_3 & 1.0 & 1.0 \\ -\lambda_3 & 0.0 & 1.0 \end{bmatrix}$$

---

[6] This assumes that $\lambda_1, \lambda_2, \lambda_3$ all exceed 0 and have values such that the model is stable.

[7] Defining $\Pi_t \equiv P_t - P_{t-1}$ as the observed inflation rate and $\Pi^*_{t+1} \equiv P^e_{t+1} - P_t$ as the expected observed inflation rate, learning rule (4) (no drift) can be expressed as $\Pi^*_{t+1} = \Pi^*_t - (1 - \lambda_1 - \lambda_2)(\Pi_t - \Pi^*_t) + (1 - \lambda_1)(\Pi_{t-1} - \Pi^*_{t-1})$. This is identical to a learning rule derived by Mussa as rational if the money supply is generated by an extrapolative-regressive process, past monetary policy is inferred by individuals, and the demand for real money balances is of the form specified by Phillip Cagan. Our estimation of system (4) in the next section can be regarded as a test of Mussa's specification.

For a vector X_0 of initial values of P^e, Π^e, and d^e, the solution of this system is

$$(10) \qquad X_t = A^t X_0 + \sum_{j=0}^{t} A^{t-j} Z_j$$

Equation (10) provides the model's predicted values of the expectation variables at every point in time as a function of an observed price series and any given set of adaptation parameters and initial conditions. For a given price series, a solution which is equivalent, but easier to implement, can be obtained by generating X_t recursively from (9). With given values of λ_1, λ_2, λ_3, and X_0, one can compute X_1 from P_1 using (9). With the value of X_1 so generated and P_2, one then computes X_2 in the same manner. By repeated use of (9) an entire time history of X_t can be simulated. It is this predicted time history of expectations which we compare with reported expectations to assess the explanatory power of the model.

II. Data and Empirical Results

Testing a theory of expectations formation is difficult because the variables to be predicted are inherently not observable. One approach is to embed the expectations equations in a model which attempts to explain other phenomena. Testing the model then involves a joint test, and it is impossible to disentangle errors in the expectations equations from errors in the remainder of the model. The Livingston survey data for the United States provide the basis for an alternative approach. Every six months since 1947 Joseph Livingston conducted a survey of economists asking their forecasts of a number of key economic variables including the Consumer Price Index (*CPI*). From the individual survey responses, Livingston constructed a "consensus" forecast which was published in late June and December. In a recent paper, John Carlson analyzed the original survey responses and generated a series for the expected *CPI*. Carlson emphasizes key timing aspects of the survey which have largely been ignored by other users of the data: namely, the information available at the time of the survey and the time horizon for which respondents were forecasting. Since timing considerations are crucial for implementation of the model, we need to examine Carlson's findings.

Consider the forecasts published in June of the *CPI* for the following December and June. To meet the publication deadline, Livingston would mail the questionnaire in mid-May along with the latest available value of the *CPI*—usually the April value. Respondents would complete the questionnaire in late May or early June. Carlson assumes that the latest information available to survey participants is the April *CPI* and that they are making eight- and fourteen-month predictions of the *CPI* to be reported for the following December and June. However, survey participants would have observed the behavior of some prices in May even though the official *CPI* had not been released. Consequently we assume that survey participants know May prices and make seven- and thirteen-month forecasts for the following December and June. For the December survey we assume that respondents know November prices and make seven- and thirteen-month forecasts for the following June and December. These timing assumptions are explored further below.

Since the survey data are available every six months, that is taken as the unit of time for equation (7). Our timing assumptions imply that P_t is the *CPI* for either May or November. At each point in time there are two observations: $(P_t^*)_7$, the average survey response for the expected *CPI* seven months hence in either December or June; and $(P_t^*)_{13}$, the average survey response for the expected *CPI* thirteen months hence in either June or December.[8] Carlson also reports standard deviations $(\sigma_t)_7$ and $(\sigma_t)_{13}$

---

[8] The Livingston survey results and the Carlson expected *CPI* data are defined with respect to four different base years. All data were converted to a base year of 1967. Carlson also presents data for nineteen-month forecasts. These data are not available for the entire time period and will not be used to estimate the model.

which indicate the dispersion of individual forecasts.[9] These standard deviations exhibit considerable variation over the data interval.

For estimation purposes we need to compute the theoretical forecasts of the expected price level for the next seven and thirteen months implied by the adaptive learning rule (7). To obtain these forecasts it is necessary to express the learning rule as a two-step procedure. Based on P_t, individuals first compute their best estimate of the price level $P_t^{e\prime}$, inflation rate $\Pi_t^{e\prime}$, and drift $d_t^{e\prime}$ which occurred for the interval $t-1$ to t. The algorithm for this stage of the learning rule is

$$\begin{aligned} &(11) \quad & P_t^{e\prime} &= P_t^e + \lambda_1(P_t - P_t^e) \\ && \Pi_t^{e\prime} &= \Pi_t^e + \lambda_2(P_t - P_t^e) \\ && d_t^{e\prime} &= d_t^e + \lambda_3(P_t - P_t^e) \end{aligned}$$

Then individuals forecast for the interval t to $t+1$ according to

$$\begin{aligned} &(12) \quad & P_{t+1}^e &= P_t^{e\prime} + \Pi_t^{e\prime} + d_t^{e\prime} \\ && \Pi_{t+1}^e &= \Pi_t^{e\prime} + d_t^{e\prime} \\ && d_{t+1}^e &= d_t^{e\prime} \end{aligned}$$

It is straightforward to verify that the two-step procedure of (11) and (12) is equivalent to (7).

The unit of time in (12) is six months while the estimation procedure requires forecasts on a monthly basis. It can be shown that the best estimate of the current monthly inflation rate is[10] $(\Pi_t^{e\prime} + 5d_t^{e\prime}/12)/6$ while the monthly drift in the current monthly inflation is $d_t^{e\prime}/36$. As a result, the theoretical forecast of the expected price level τ months in the future is given by

$$(13) \quad (P_t^e)_\tau = P_t^{e\prime} + \frac{\tau}{6}\left(\Pi_t^{e\prime} + \frac{5}{12}d_t^{e\prime}\right) + \frac{\tau(\tau+1)}{2}\,\frac{d_t^{e\prime}}{36}$$

According to (13), the seven- and thirteen-month price forecasts implied by the adaptive learning rule are $(P_t^e)_7 = P_t^{e\prime} + 7\Pi_t^{e\prime}/6 + 91d_t^{e\prime}/72$ and $(P_t^e)_{13} = P_t^{e\prime} + 13\Pi_t^{e\prime}/6 + 247d_t^{e\prime}/72$, respectively.

For a given time history of the *CPI*, the model solution of (10) and the theoretical price forecast of (13) depend only on the initial conditions P_0^e, Π_0^e, and d_0^e and on the adaptation coefficients λ_1, λ_2, and λ_3. The ability of the model to reproduce the survey price forecasts is indicated by the weighted residuals $\Delta s_{t,1} = [(P_t^*)_7 - (P_t^e)_7]/w_{t,1}$ and $\Delta s_{t,2} = [(P_t^*)_{13} - (P_t^e)_{13}]/w_{t,2}$ where the weights $w_{t,1}$ and $w_{t,2}$ are computed from the standard deviations of the price forecasts.[11]

---

[9] Since the model is estimated in logarithmic form, the standard deviations reported by Carlson were divided by the price level to obtain an approximate standard deviation for the logarithm of the expected price.

[10] π_t denotes the change in underlying prices over the most recent six months and d_t its change from the preceding six-month period. Let β denote the change in the price level over the most recent month, or current monthly inflation rate. Assuming that the monthly inflation rate drifted the same amount in each of the past six months, the monthly change in the six-month inflation rate was $d_t/6$ and the monthly change in the monthly inflation rate was $d_t/36$. Thus if the most recent monthly inflation rate was β, then the previous month's was $\beta - d_t/36$, the next previous $\beta - 2d_t/36$, and so on. The sum of the preceding six months' price changes is $6\beta - 5d_t/12$ and must equal π_t. Solving for β yields $\beta = (\pi_t + 5d_t/12)/6$. Substituting the current estimates $\Pi_t^{e\prime}$ and $d_t^{e\prime}$ for π_t and d_t, respectively, gives the claimed result.

[11] The standard deviations of the expected price data reported by Carlson exhibit considerable variation over the data interval. This creates a potential problem of heteroscedasticity because the poorer quality observations may have larger residuals and dominate the estimation process. If this were the only source of error, the weights could simply be set equal to the standard deviations of the observations. However, if other errors exist, the weights should be a function of the standard deviations of both sources of error. To allow for other errors we have used the following simple weighting scheme. We first compute the average standard deviations $(\sigma^a)_7$ and $(\sigma^a)_{13}$ and then compute the normalized standard deviations $(\sigma_t^*)_7 = (\sigma_t)_7/(\sigma^a)_7$ and $(\sigma_t^*)_{13} = (\sigma_t)_{13}/(\sigma^a)_{13}$. Finally, the weights are computed as $w_{t,1} = 1 - \alpha + \alpha(\sigma_t^*)_7$ and $w_{t,2} = 1 - \alpha + \alpha(\sigma_t^*)_{13}$ where α is a constant between zero and one. If $\alpha = 1$ the weights equal the standard deviations (normalized to give an average weight of unity) while if $\alpha = 0$ all observations receive unit weight. The final value of α will be selected to give the minimum weighted residual variance.

TABLE 1

Estimate	P_0^e	Π_0^e	d_0^e	λ_1	λ_2	λ_3	ρ	*D.W.*	*SSR*	*F*-Test
1947–75[a] No Constraints	4.154	−0.014	−0.0005	1.236	0.063	0.002	0.462	1.956	0.0043	–
1947–75 $\lambda_3 = d_0^e = 0.0$	4.161	−0.023	–	1.225	0.096	–	0.519	2.001	0.0044	1.267
1974–75 $\lambda_1 = 1.0$	–	−0.023	−0.0004	–	0.099	0.001	0.540	1.936	0.0056	16.477
1947–61 No Constraints	4.152	−0.010	0.0004	1.355	0.012	0.0004	0.509	1.755	0.0030	1.294
1961–75 No Constraints	4.500	0.005	0.0001	1.095	0.095	0.0003	0.367	2.068	0.0009	
April/October Prices 8- and 14-month forecast	4.141	−0.006	−0.0006	1.187	0.040	0.003	0.589	1.827	0.0049	15.070
June/December Prices 6- and 12-month forecast	4.154	−0.017	−0.0006	1.108	0.067	0.002	0.243	2.125	0.0058	37.647
April/October Prices 6- and 12-month forecast	4.145	−0.012	−0.0009	1.174	0.069	0.004	0.648	1.816	0.0050	17.581
May/November Prices 6- and 12-month forecast	4.153	−0.015	−0.0008	1.237	0.074	0.003	0.455	1.959	0.0043	~0

[a] The first five rows assume seven- and thirteen-month forecasts updated on May/November prices.

The estimation entails finding those values of the adaptation parameters and initial conditions which minimize the sum of squared transformed residuals (*SSR*)

$$SSR = \sum_{t=1}^{T} \sum_{j=1}^{2} (\Delta s_{t,j} - \rho \Delta s_{t-1,j})^2$$

where ρ is the first-order serial correlation coefficient.[12] Since the parameter estimates are non-linear in the initial conditions, adaptation coefficients and ρ, the minimization of *SSR* requires an iterative procedure. We use the technique of Donald Marquardt based on interpolation between the Taylor series and gradient methods.[13]

The first row of Table 1 presents estimation results for the unrestricted model.[14] The initial conditions correspond to negative expected inflation rate with negative expected drift. Individuals expected prices to fall in the immediate postwar period. The negative expected inflation rates did not materialize and Π_t^e and d_t^e turned positive as individuals adapted to the actual price behavior. The second row of Table 1 tests whether the time varying drift was important in explaining the survey data. The *F*-test cannot reject the hypothesis of zero drift at the 95 percent level of significance.[15]

Was the provision for beliefs in purely transitory shocks to the price level important? The third row of Table 1 provides estimates of the model specified by (8). That specification requires that λ_1 equals

---

[12] If the actual error term has first-order serial correlation then the estimation procedure is equivalent to obtaining maximum likelihood estimates. The only difference is the term $log(1-\rho^2)$ in the likelihood function which would have a negligible impact on our estimates.

[13] The model was iterated to convergence starting with widely different initial parameter estimates to insure that we had obtained a global maximum.

[14] The model was estimated using values of α in the weighting scheme which varied from 0 to 1 in increments of 0.25. The value $\alpha = 0.75$ gave the minimum *SSR* and is used in all the estimates discussed below. This value of α implies that the weights were primarily a function of the observation standard deviations. The estimation results are not overly sensitive to the weighting scheme employed. For example, reducing α to 0.25 so that the weights are nearly unity only increases *SSR* from 0.0043 to 0.0046. The parameters obtained in this case are $\lambda_1 = 1.341$, $\lambda_2 = 0.102$, and $\lambda_3 = 0.0005$.

[15] We assume that the test statistic $V = (T - K/q) \times (SSR^+ - SSR/SSR)$, where SSR^+ is the constrained *SSR*, *T* is the number observations, *K* is the number of fitting parameters, and *q* is the number of constraints, is distributed $F(q, T-K)$. The R^2 is computed in the usual fashion as $1 - SSR/SST$ in which *SST* is defined as the minimum sum of squared weighted residuals possible for a constant model prediction.

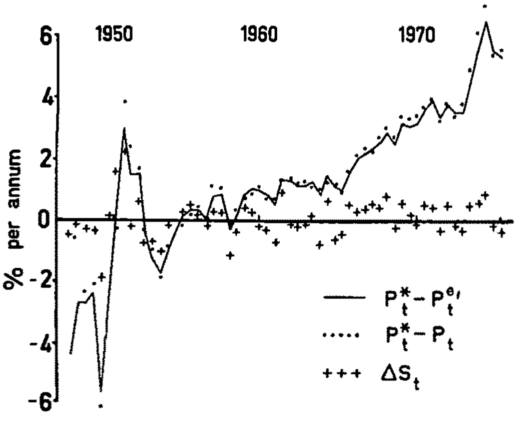


FIGURE 1. EXPECTED UNDERLYING INFLATION, EXPECTED OBSERVED INFLATION AND MODEL RESIDUALS FOR THIRTEEN-MONTH FORECASTS

unity—individuals form their expectations on the basis of observed inflation rates. The *F*-test rejects this hypothesis at the 99 percent level of significance against the alternative that λ_1 exceeds unity. Moreover, the fact that $\lambda_1+\lambda_2+\lambda_3$ significantly exceeds unity appears to rule out the short-run regressivity of expected observed inflation conjectured by Frenkel and Mussa. Many users of Livingston's survey data transform the price level forecasts into expected inflation rates by subtracting the current price level and converting to a growth rate. This transformation is suspect because it ignores the potentially substantial difference between P_t and $P^{e'}_t$ when λ_1 differs from unity. Subtracting P_t as opposed to $P^{e'}_t$ yields the expected observed inflation rate as opposed to the expected underlying inflation rate. Figure 1 depicts these two series for the thirteen-month forecasts. Also shown in Figure 1 is the residual between the expected underlying inflation rate implied by the data and that generated by the adaptive algorithm.[16]

Stephen Turnovsky, in a previous analysis of the Livingston data, concluded that the process of expectations formation was fundamentally different between the period after 1960 and the earlier period. The fourth and fifth rows of Table 1 contain separate estimates of the model for the 1947–61 and 1961–75 subperiods, respectively. At first glance the estimation results appear to support Turnovsky. Quite different values of λ_1 and λ_2 are obtained for the two data intervals (in both cases the drift has a negligible impact on *SSR*). However, these parameters have large standard errors for the shorter samples and the *F*-test indicates that the parameters do not differ significantly at the 95 percent level between the two periods. We find no compelling evidence of a change in the process of expectations formation.

The remaining model estimates presented in Table 1 examine timing questions raised by Carlson. The sixth and seventh rows consider the information available to survey participants. Row six embodies Carlson's assumption of eight- and fourteen-month forecasts based on April and October prices, while row seven assumes six- and twelve-month forecasts based on June and December prices. Both hypotheses lead to a significant increase in *SSR* over that obtained in the first row. The *F*-tests reject both timing hypotheses at the 99 percent level of significance in favor of seven- and thirteen-month forecasts based on May and November prices.[17]

The final estimates address the question of forecast horizon. In a separate analysis of the Livingston data, Paul Wachtel assumed that individuals knew April and October prices but were making forecasts of the price level only six and twelve months into the future rather than the eight and fourteen months assumed by Carlson. As Carlson notes, survey responses consistently underpredict the subsequently realized *CPI*. For example, the eight-month forecast underpredicts the ultimately realized *CPI* for fifty of the fifty-eight observations. If the forecast horizon is shortened to six months then only thirty-six of the forecasts underpredict the subsequently realized *CPI*. The eighth row of Table 1 presents model estimates under Wachtel's timing assumptions. There is an insignificant increase in *SSR* compared to the estimates in the sixth row based on Carlson's timing assumptions. However, the adaptation coefficients λ_2 and λ_3 are significantly larger. Since the shorter forecast horizon eliminates the consistent prediction errors, larger values of λ_2 and λ_3 are necessary to generate the required variations in Π_t^e and d_t^e. The *F*-test indicates that this timing hypothesis is also rejected in favor of the first row.

---

[16] The annualized expected underlying inflation, denoted as $P^*_t - P^{e'}_t$ in Figure 1, is computed as $[ln(P^*_t)_{13} - ln(P^{e'}_t)]\cdot 1200/13$ while the expected observed inflation rate, denoted as $P^*_t - P_t$, is $[ln(P^*_t)_{13} - ln(P_t)]\cdot 1200/13$. There were no values of R^2 listed in Table 1 because the large changes in P_t over the data set make it a very optimistic indicator of the ability of the model to reproduce the expectations data (the value of R^2 for the first row estimate is 0.99998). The model residuals compared to the expected inflation rate in Figure 1 is a tougher test of the goodness of fit. An R^2 computed on this basis has a value of 0.89.

[17] This conclusion is conditioned, of course, on the validity of the expectations adjustment specification.

The last row of Table 1 considers six- and twelve-month forecasts based on May and November prices. In terms of goodness of fit, this timing hypothesis is indistinguishable from seven- and thirteen-month forecasts based on the same prices. The shorter forecasting horizon in this instance reduces the size of the consistent prediction errors inherent in the data. Larger estimated values of λ_2 and λ_3 are obtained as a result.

III. Conclusion

The multilevel adaptive scheme replicates the expectations adjustment process closely. Simulation of the model explains 0.89 of the variation in expected underlying inflation rates implied by the survey data for thirteen-month forecasts, with two-thirds of the residuals lying under 0.5 percent per year. This "goodness of fit" is all the more surprising since the only input is past observed prices; no data on previous reported expectations, which might have served as proxies for omitted variables, is used.

Within the context of the model, the provision for expectations about trends in inflation rates added little explanatory power. But the provision for distinguishing transitory from permanent shifts in the price level, or observed from underlying prices, proved important. The adaptation coefficients displayed remarkable stability over long periods of quite different inflation experience.

A test of the type conducted in this paper cannot resolve the issue of whether or not inflation expectations are generated rationally on the basis of information other than observed prices. What it does indicate is that reported expectations behave to a large extent as if formulated on the basis of past prices alone, and that an individual using a naive algorithm could have generated much the same expectations at very little cost.

REFERENCES

P. Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956.

J. A. Carlson, "A Study of Price Forecasts," *Annals Econ. Soc. Measure.*, Winter 1977, *6*, 27–56.

E. L. Feige and D. K. Pearce, "Economically Rational Expectations," *J. Polit. Econ.*, June 1976, *84*, 499–522.

J. A. Frenkel, "Inflation and the Formation of Expectations," *J. Monet. Econ.*, June 1975, *1*, 403–21.

R. L. Jacobs and R. A. Jones, "A Bayesian Approach to Adaptive Expectations," disc. paper no. 93, Univ. California-Los Angeles, June 1977.

Joseph A. Livingston, surveys published twice yearly, *Philadelphia Sunday Bulletin*, 1948–71; *Philadelphia Inquirer*, 1972 on.

D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Non-Linear Parameters," *SIAM J. Appl. Math.*, June 1963, *11*, 431–41.

M. Mussa, "Adaptive and Regressive Expectations in a Rational Model of the Inflationary Process," *J. Monet. Econ.*, June 1975, *1*, 423–42.

J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts," *J. Amer. Statist. Assn.* June 1960, *55*, 299–306.

M. Nerlove and S. Wage, "On the Optimality of Adaptive Forecasting," *Manage. Sci.*, Jan. 1964, *10*, 207–24.

S. J. Turnovsky, "Empirical Evidence on the Formation of Price Expectations," *J. Amer. Statist. Assn.*, Dec. 1970, *65*, 1441–54.

P. Wachtel, "Survey Measures of Expected Inflation and their Potential Usefulness," in Joel Popkin, ed., *Analysis of Inflation 1965–1974*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 42, Cambridge 1977.

Multinational Firms and the Theory of International Trade and Investment

By RAVEENDRA N. BATRA AND RAMA RAMACHANDRAN*

Despite increased importance of the multinational corporation in the arena of economic activity and political influence, trade theorists have either ignored it or expressed unguarded skepticism at the ability of the conventional trade models to successfully capture and analyze the features of this new phenomenon.[1] The purpose of this paper is twofold: first, by building upon the contribution by Richard Caves, we will show that the traditional trade models can be adapted in a way that preserves most, if not all, of the attributes introduced by international firms.[2] Second, we will conduct a comparative statics analysis to explore the implications of tariffs and taxes for resource allocation and international capital movements. Our results here confirm what has already been well established in myriad empirical studies, that maturation of the international firm has vastly increased economic interdependence among trading countries, and that few nations can eschew the ripples caused by economic policies of other nations.[3]

I. Assumptions and the Model

The conventional general equilibrium models of international trade and investment, expounded in detail by Murray Kemp and Ronald Jones (1967), rely heavily on the properties of the Heckscher-Ohlin model. This model is ill-equipped to tackle the specificity of some factors and technology which we consider to be one of the main characteristics of international investments by multinational firms. Almost invariably the multinational firm possesses oligopolistic control over some patents and specific technology which enables it to outcompete local firms in foreign countries. The return to such investments includes not only an adequate return on its capital but also on its inventions, managerial know-how, superior marketing techniques, etc., which it has already developed. While it might have incurred considerable expenses in developing this know-how, the opportunity cost of transmitting such inventions abroad is practically zero, and whatever royalties and fees that are thus received are pure profits.

In this paper we intend to capture these features of the multinational firm by analyzing international investment using the specific factor model of international trade which was popular before 1944 when the contribution by Wolfgang Stolper and Paul Samuelson appeared, and which was recently resurrected by Batra and Francisco Casas, among others. We present below a two-sector model where the multinational capital is specific to one sector while the other sector consists of national and local firms with production activities confined to one nation. The multinational capital, while specific to one sector within each nation, is nevertheless mobile between the same sectors in different nations.[4]

---

*Professor and assistant professor, respectively, Southern Methodist University. We are indebted to George Borts and an anonymous referee for many helpful comments.

[1] See for instance Thomas Horst, and Robert Baldwin and J. David Richardson, p. 264, for expressions of similar opinion.

[2] Actually Caves investigated the implications of the emergence of the international firms in terms of both partial and general equilibrium models. Our analysis builds further upon his general equilibrium system.

[3] See Richard Barnett and Ronald Muller for an eloquent presentation of this view.

[4] See Caves, p. 17. The argument of our paper is based on an assumption that (i) the multinational and local sectors are distinctly defined, and (ii) the multinational sector is a significant part of economic activity in the host country. Regarding (i) there is no problem: Jean Servan-Schreiber points out that in all Europe, the *U.S.*-based firms by the late 1960's produced 80 percent of computers, 95 percent of integrated circuits, 50

The ownership of patents, the availability of special managerial know-how, and the marketing skills are represented by a third factor S in the production function of the multinational firm. Some properties of the specific factor S are noteworthy. Following Harry Johnson, we shall assume that S is measured not in terms of a physical unit but in terms of some intangible abstract unit whose transference by international firms into the host country does not diminish its use in the source country.[5] It is assumed that S is a "state variable" whose value at any instant of time depends on the past efforts at developing the specific skills possessed by the multinational firm. It may be interpreted as an index of Harvey Leibenstein's X-efficiency factor; Leibenstein (1966, p. 412) mentions the nonmarketability of certain inputs as a factor responsible for differing X-efficiency of firms, which is exactly our assumption regarding the role of S. As argued by Leibenstein (1966, 1978) and T. Y. Shen, X-efficiency leads to dispersion in the level of productivity of firms at any instant of time. We shall, therefore, assume that $0<S<\bar{S}$, where $\bar{S}$ is the saturation level at which the marginal product of S is zero. One particular merit of this approach is that the return to the intangible factor S is a pure surplus and captures, in our model, any excess profit earned by multinational firms. In this way, one could incorporate all the features of a world dominated by global firms (excess profits, industry-specific capital and technology, etc.) in a general equilibrium model and still retain the assumptions of perfect competition that considerably simplifies exposition and, in most cases, is the only way in which a general equilibrium analysis can be carried out.[6]

The model developed below will then be based on the following assumptions:

ASSUMPTION 1: *There are two countries, a source country (the home of the multinational firm) and a host country (which is foreign to the global firm); and two sectors, one wherein the multinational firms compete with each other and the other consisting of goods produced by the local firms.*

ASSUMPTION 2: *Labor is the only nonspecific factor and capital is specific to both sectors; the production function of the multinational firms also contains another specific factor that represents patents, technical and managerial know-how, etc.*

ASSUMPTION 3: *All other assumptions of the Heckscher-Ohlin model are retained. Thus perfect competition, linearly homogeneous and concave production functions, full employment, and inelastic factor supplies are also assumed.*

---

percent of semiconductors, and so on. The *U.S.* direct foreign investment is thus concentrated mostly in advanced technology, capital-intensive industries. As regards (ii), the multinational firms perhaps account for less than 10 percent of economic activity in all Europe. The important point, however, is that their assets are growing at phenomenal rates; Barnett and Muller, p. 26, report that by some estimates the direct foreign investment is likely to reach $4 trillion by 1985, or about 54–80 percent of all production assets of the noncommunist countries. Thus, the productive activity of the multinational firms is expected to become a colossus in many host countries; therefore the implications of a distinctive multinational sector is worth exploring. We are also ignoring the phenomena of cross hauling of international investment wherein firms in each country invest and control production facilities in the other country. Despite some recent inroads made by non-American multinational firms, the *U.S.*-based global corporations were by 1969 the biggest suppliers of direct foreign investment. Also a significant portion of direct foreign investment in the United States is in the form of export subsidiary, one that facilitates importing goods, while *U.S.*-based firms set up production facilities abroad. See Sidney Rolfe for further details.

[5] S may be measured in terms of man-hours spent in the conduct of research or in terms of financial capital used in the discovery of patents and so on. Johnson emphasized the point that the know-how and managerial skills have the nature of public goods which have high fixed costs but low marginal costs. C. F. Pratten observes: "From my discussion with officials of companies it was clear that management of national operations of most companies have access to some knowledge of production and marketing techniques, and they can obtain advice from other companies in the same group" (pp. 50–51).

[6] It seems to us that in order to investigate what the emergence of multinational firms means to the world economy, the general equilibrium analysis—as far as it goes—is more appropriate than the partial analysis.

ASSUMPTION 4: *Both the countries are small and take as given the relative prices that are determined in the rest of the world.*[7]

Let X and Y, respectively, be the sectors in which the multinational and the local firms produce their products in the source country. The two aggregate production functions are

$$(1)\quad X=X(L_x,K_x,S)\qquad X^*=X^*(L_x^*,K_x^*,S^*)$$

$$(2)\quad Y=Y(L_y,K_y)\qquad Y^*=Y^*(L_y^*,K_y^*)$$

where L stands for labor, K for capital, S for a specific factor, and the star denotes the variables in the host country. The production functions, which are assumed to be linearly homogeneous and concave, satisfy the following properties: In X

$$(3a)\qquad X_j>0,\quad X_{jj}<0,\quad X_{jk}>0\qquad (j,k=K,L,S;j\neq k)$$

$$X_{LL}X_{KK}-X_{KL}^2>0$$

and in Y

$$(3b)\quad Y_j>0,\quad Y_{jj}<0,\quad Y_{KL}>0\quad (j=L,K)$$

$$Y_{LL}Y_{KK}-Y_{KL}^2=0$$

where X_j is the marginal product of the jth factor, X_{jj} is the change in the marginal product of the jth factor with respect to its own input (i.e., $\partial X_L/\partial L_x=X_{LL}$), and X_{jk} is the change in the marginal product of the jth factor with respect to the kth input ($j\neq k$); similar properties characterize the function $Y(\cdot)$.

With regard to capital, we assume that K_x and K_y are also specific to each sector; labor is a nonspecific factor and is mobile between two sectors, so that only one wage rate prevails within each country.

The firm wants to maximize global profits after taxes. Generally both the source country and the host country levy corporate income tax. We shall assume, in accordance with current *U.S.* tax practice, that the source country gives limited tax credit for the tax payed in the host country; the maximum tax credit that the corporation can earn is estimated by applying the source-country tax rate to the income earned in the host country. Further assume that the goods produced by the multinationals in the host country are importable goods there and that the host country levies a tariff on its imports. Hence $p^*=p(1+\tau^*)$ where $\tau^*\geqslant 0$ is the host country's tariff rate on its import of good X.

The relative price of X in terms of Y, p is given by the world market. The firm takes the wage (w) and the rental (r) rate in both countries as given. Also let F be the expenses incurred by the firm for research and managerial development necessary to maintain its global leadership. The global profits may be computed as follows:

Tax base in host country $=B^*$

$$=(p^*X^*-w^*L_x^*)$$

Tax base in source country $=B$

$$=(pX-wL_x+p^*X^*-w^*L_x^*-F)$$

Tax credit[8] in source country $=C$

$$=(p^*X^*-w^*L_x^*)t$$

Global profit after taxes[9] is given by

$$\begin{aligned}(4)\quad \Pi &= B(1-t)-B^*t^*+C-r^*K_x^*-rK_x\\ &=(pX-wL_x-F)(1-t)\\ &\quad+(p^*X^*-w^*L_x^*)(1-t^*)-rK_x-r^*K_x^*\end{aligned}$$

---

[7] This assumption is reminiscent of Jacob Viner's well-known customs union model where two or three small countries trade with each other at given terms of trade that are determined in the rest of the world.

[8] Batra and Josef Hadar point out that if $t^*<t$, then t^* does not influence the decision of the multinational firm, as C now equals $(p^*X^*-w^*L_x^*)t^*=B^*t^*$.

[9] Actually, following Horst, we should also include the intrafirm exports (or imports), i.e., the exports (or imports) from the parent firm to the subsidiary, in the global profit equation. However, such trade is implicit in our definition. For example, suppose M is the quantity exported by the multinational firms to their subsidiaries in the host country and D_x is the quantity

The marginal (first-order) conditions for profit maximization by the multinational firm require that

$$pX_L = w_x \tag{5}$$

$$p^*X_L^* = w_x^*$$

$$pX_k(1-t) = r_x \tag{6}$$

$$p^*X_K^*(1-t^*) = r_x^*$$

Similarly, profit maximization by local firms leads to

$$Y_L = w_y \quad \text{and} \quad Y_L^* = w_y^* \tag{7}$$

$$Y_K = r_y \quad \text{and} \quad Y_K^* = r_y^* \tag{8}$$

Mobility of labor in each country ensures that the wage rate is the same in both sectors, so that

$$w = w_x = w_y \quad \text{and} \quad w^* = w_x^* = w_y^* \tag{9}$$

The rental of capital need not be the same within each country because of limited mobility and industry specificity, but capital stock owned by the international firms is mobile from one country to the other, and this mobility generates an equality between r_x and r_x^*. Hence

$$r_x = r_x^* \tag{10}$$

From equations (5)–(10) we can obtain the factor-market equilibrium conditions which determine the allocation of labor between the two sectors in each country and of the multinational capital between the two countries. From (5), (7), and (9) we have

$$pX_L(L_x, K_x, S) = Y_L(L_y, K_y) \tag{11}$$

$$p^*X_L^*(L_x^*, K_x^*, S^*) = Y_L^*(L_y^*, K_y^*) \tag{12}$$

These equations determine the allocation of labor between the two sectors in each country.

From (6) and (10),

$$p(1-t)X_K(L_x, K_x, S) = p^*(1-t^*)X_K^*(L_x^*, K_x^*, S^*) \tag{13}$$

This equation determines the allocation of the multinational capital between the source and the host country.

From full employment of labor in each country,

$$L_x + L_y = \bar{L} \tag{14}$$

$$L_x^* + L_y^* = \bar{L}^* \tag{15}$$

and from the constraint imposed by the stock of capital owned by the multinational firms,

$$K_x + K_x^* = \bar{K}_x \tag{16}$$

where $\bar{L}$, $\bar{L}^*$, and $\bar{K}_x$ are constant. Similarly, since all factors are in inelastic supply, K_y and K_y^* are also constant.

Thus we are able to incorporate the basic attributes of multinational firms in a general equilibrium analysis and use it below to undertake comparative static analysis of the impact of changes in taxes and tariffs. Resource allocation within each country and

---

demanded of X, then $M = D_x^* - X^* = X - D_x$. In view of this, global profits (ignoring taxes) become

$$\Pi = (pD_x + pM - w_xL_x - r_xK_x) + (p^*D_x^* - p^*M - w_x^*L_x^* - r_x^*K_x^*) - F$$

which is similar to Horst's definition. Since we have assumed that the production function is linear homogeneous, it is possible to define implicitly (from the product exhaustion rule) the marginal product of the intangible factor, X_S and X_S^*,

$$X = X_LL_x + X_KK_x + X_SS$$

$$X^* = X_L^*L_x^* + X_K^*K_x^* + X_SS^*$$

Substituting equations (5)–(8) into (4) would then show that

$$\Pi = p(1-t)X_SS + p^*(1-t^*)X_S^*S^* - (1-t)F$$

The second term represents the addition to pure surplus of the multinational firm from utilizing its expertise for production abroad.

$$(18)\quad \begin{bmatrix} p^*X_{LL}^*+Y_{LL}^* & p^*X_{LK}^* & 0 \\ 0 & -pX_{LK} & pX_{LL}+Y_{LL} \\ p^*(1-t^*)X_{LK}^* & \left\{\begin{matrix} p^*(1-t^*)X_{KK}^* \\ +p(1-t)X_{KK}\end{matrix}\right\} & -p(1-t)X_{KL} \end{bmatrix} \begin{bmatrix} dL_x^* \\ dK_x^* \\ dL_x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \left\{\begin{matrix} p^*X_K^*dt^* \\ -pX_Kdt \end{matrix}\right\} \end{bmatrix}$$

the process of capital movements between the two countries can be determined with the help of six equations given by (11)–(16), containing six variables (L_x, L_y, L_x^*, L_y^*, K_x, and K_x^*) and eleven parameters ($\bar{S}$, $\bar{S}^*$, $\bar{K}_y$, $\bar{K}_y^*$, $\bar{L}$, $\bar{L}^*$, $\bar{K}_x$, p, t, t^*, and τ^*).

From our assumption that the multinational firms operate in small countries and sell their goods in those countries as well as the rest of the world, the international terms of trade (p) can be treated as constant. And with a given τ^*, p^* is also constant. At the present state of the literature dealing with the implications of the multinational firms, this seems to be an appropriate assumption, although a complete analysis should also examine the case where the terms of trade are variable.

II. Taxes on International Investment

The model can be used to analyze the effects of a change in the corporate income tax. Substituting for L_y, L_y^*, K_x from (14) to (16) into (11), (12), and (13), we get[10]

$$(17)\quad p^*X_L^*(L_x^*,K_x^*,S^*)=Y_L^*(\bar{L}^*-L_x^*,K_y^*)$$

$$pX_L(L_x,\bar{K}_x-K_x^*,S)=Y_L(\bar{L}-L_x,K_y)$$

$$p^*(1-t^*)X_K^*(L_x^*,K_x^*,S^*)$$

$$=p(1-t)X_K(L_x,\bar{K}_x-K_x^*,S)$$

With these substitutions, our model is considerably simplified. The system of equations (17) contains three equations and three unknowns, L_x^*, K_x^*, and L_x. Differentiating these totally we obtain equation (18) shown above, or

$$(18')\quad [A][B]=[C]$$

where $[A],[B],[C]$ are the three matrices of (18). The determinant of (18) is given by

$$(19)\quad D=-\Big[(p^*X_{LL}^*+Y_{LL}^*)p^2(1-t)$$

$$\times\left(H_x+\frac{1}{p}X_{KK}Y_{LL}\right)$$

$$+(pX_{LL}+Y_{LL})p^{*2}(1-t^*)$$

$$\times\left(H_x^*+\frac{1}{p^*}X_{KK}^*Y_{LL}^*\right)\Big]$$

where $H_x=X_{LL}X_{KK}-X_{KL}^2>0$, and similarly for H_x^*. In view of our assumptions about the production function, D is positive.

Let us begin our analysis of the income tax on the multinational investment by examining the response of the variables in the host country to a change in the tax rate in that country. With $dt=0$, the solution of (18) yields

$$(20)\quad \frac{dL_x^*}{dt^*}=\{p^{*2}X_K^*X_{KL}^*(pX_{LL}+Y_{LL})\}$$

$$\div D=-\frac{dL_y^*}{dt^*}$$

$$(21)\quad \frac{dK_x^*}{dt^*}=-\{p^*X_K^*(p^*X_{LL}^*+Y_{LL}^*)$$

$$(pX_{LL}+Y_{LL})\}\div D=-\frac{dK_x}{dt^*}$$

Clearly $dL_x^*/dt^*<0$ and $dK_x^*/dt^*<0$, and we obtain an eminently plausible result that the international investment tax leads to a decline in the employment of the multinational capital in the host country. Since $d\bar{K}_x$

---

[10] If $\tau^*=0$, then $p=p^*$.

$=0$, this obviously implies that capital moves back to the source country. As the host country imposes a tax, investment there becomes less attractive because the marginal product of capital net of the tax declines; capital movement back to the source country occurs until the net marginal product of capital in the host country rises to that of capital employed in the source country.

With capital so moving, the marginal product of labor declines in X^* (because $X^*_{KL} > 0$), so that labor moves towards the goods produced by local firms in the host country. That is why $dL^*_x/dt^* < 0$ and hence $dL^*_y/dt^* = -dL^*_x/dt^* > 0$. It also follows that the output of Y^* rises and that of X^* declines.

What happens to factor rewards in the host country? Since $w^* = Y^*_L(L^*_y, K^*_y)$, and K^*_y is constant at any moment of time,

$$\frac{dw^*}{dt^*} = Y^*_{LL}\frac{dL^*_y}{dt^*} < 0$$

so that the real wage rate declines.[11] The reason is that as the marginal product of labor declines in goods produced by the multinationals, demand for labor in X^* declines, creating excess supply of labor in the entire economy, and this forces the real wage to go down to clear the labor market.

This result gains further comprehension from Figure 1 where the real wage rate expressed in terms of Y^* is measured along the vertical axis and labor along the horizontal axis: BC is the negatively sloped demand curve for labor in X^* and AD is the aggregate demand curve for labor, where AD is obtained by a horizontal addition of the demand for labor in Y^* to that demand in X^* at any w^*. For simplicity these curves are drawn to be linear. The inelastic supply curve of labor is given by $R\bar{L}^*$, the equilibrium wage by w^*_e, labor employed in X^* by OG, and that employed in Y^* by GR. This is how the initial situation is depicted. Now suppose a tax on the multinational investment is imposed by the host country and K^*_x declines; this causes a downward shift of the labor demand curve in X^* to $B'C'$ and of the aggregate labor demand curve to $A'D'$ such that the distance MQ equals the distance HS. The equilibrium wage rate declines to w^*_t, the employment of labor in X^* declines to OG' and that in Y^* rises to $G'R$. A similar diagram can be used to depict the flight of the multinational capital from the host country.

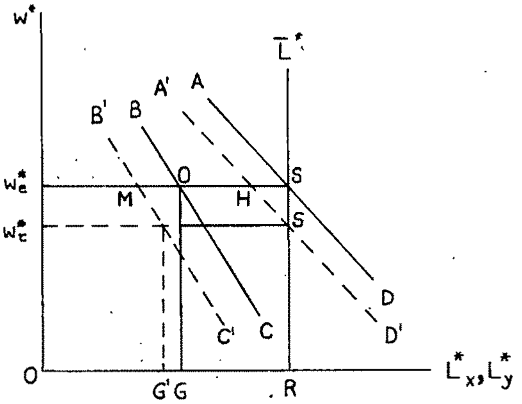


FIGURE 1

What about the return to capital? We know that the equilibrium $r^*_x = r_x = p(1-t)X_K(L_x, K_x, S)$. Therefore,

(22)

$$\frac{dr_x}{dt^*} = p(1-t)X_{LK}\frac{dL_x}{dt^*} + p(1-t)X_{KK}\frac{dK_x}{dt^*}$$

From (18),

$$\frac{dL_x}{dt^*} = -p^*pX^*_KX_{LK}(p^*X^*_{LL} + Y^*_{LL}) \div D > 0$$

Substitute this relation and (21) into (22) to obtain

$$\frac{dr_x}{dt^*} = p^2p^*X^*_K(p^*X^*_{LL} + Y^*_{LL})$$

$$\cdot\left(H_x + \frac{1}{p}Y_{LL}X_{KK}\right) \div D < 0$$

Hence the rental on the multinational

---

[11] The same result can be obtained by differentiating

$$w^* = X^*_L(L^*_x, K^*_x, S^*)$$

capital declines in both countries. The rentals on the local capital respond differently. For

$$\frac{dr_y^*}{dt^*} = Y_{KL}^* \frac{dL_y^*}{dt^*} > 0$$

Whereas $$\frac{dr_y}{dt^*} = Y_{KL} \frac{dL_y}{dt^*} < 0$$

Intuitively, the decline of employment in the multinational firm in the host country leads to a movement of labor to the local sector; this increases the rent of capital in that sector as capital is not free to move between sectors. The fall in demand for labor also leads to a fall in wages. An opposite effect occurs in the source country. The transfer of capital to the source country by the multinational firm leads to an increase in the wage rate and a decline in the rent of local capital in that country.

Consider now the effects of a change in the tax rate on the total profits earned by the multinational firm. Differentiating (4), it can be easily shown that

$$(23)\quad \frac{d\Pi}{dt^*} = \underset{(+)}{(p^*X^* - w^*L_x^*)} - \underset{(+)}{(1-t)L_x \frac{dw}{dt^*}} - \underset{(-)}{(1-t^*)L_x^* \frac{dw^*}{dt^*}} - \underset{(-)}{\bar{K}_x \frac{dr_x}{dt^*}}$$

The sign of each of the four terms is shown in parentheses below. The first term is the negative of the total income earned by multinationals in the host country. The second and third terms refer to change in wage payments in the two, assuming that employment is held constant; the two terms have opposite signs. The last term refers to the change in rent earned by the capital of the multinational firm. It is very likely that the first term determines the sign of the expression as the other expressions partially cancel each other; hence an increase in taxes is likely to reduce global profits. If we compute the change in total income (profits plus rent on its capital) earned by multinationals

$$\frac{d\Pi}{dt^*} + \bar{K}_x \frac{dr_x}{dt^*}$$

then it is all the more negative. Thus, the opposition of the multinationals to corporate income taxes is not surprising.

Let us now consider how the tax affects the national income (I^*) of the host country:

(24)

$$\begin{aligned} I^* &= p^*X^* + Y^* - (1-t^*)(p^*X^* - w^*L_x^*) \\ &= p^*X^* + Y^* - (1-t^*)\{p^*X_K^*K_x^* \\ &\quad + (p^*X^* - p^*X_L^*L_x^* - p^*X_K^*K_x^*)\} \end{aligned}$$

where $(1-t^*)(p^*X^* - w^*L_x^*)$ is the income which the multinationals repatriate to the source country. Differentiating (24) and using (6), we get

$$\begin{aligned} (25)\quad \frac{dI^*}{dt^*} &= \underset{(+)}{\left(p^*X^* - p^*X_L^*L_x^* - p^*X_K^*K_x^*\right)} \\ &\quad + t^*p^*X_K^* \underset{(-)}{\frac{dK_X^*}{dt^*}} - K_x^* \underset{(-)}{\frac{dr_x^*}{dt^*}} \\ &\quad + (1-t^*)\left[p^*(X_{LL}^*L_x^* + X_{LK}^*K_x^*) \underset{(-)}{\frac{dL_x^*}{dt^*}} \right. \\ &\quad \left. + p^*(X_{KK}^*K_x^* + X_{LK}^*L_x^*) \underset{(-)}{\frac{dK_x^*}{dt^*}} \right] \end{aligned}$$

Given equation (3),[12] $(X_{LL}^*L_x^* + X_{LK}^*K_x^*)$ and $(X_{KK}^*K_x^* + X_{LK}^*L_x^*)$ are negative. The signs of the other expressions are noted below them in equation (25). $t^*p^*(dK_x^*/dt^*)$ is the only negative term in the equation. If t^* is small, or if initially the host country has no tax so that $t^*=0$, then the whole expression is unambiguously positive, showing that an introduction of the corporate income tax by the host country raises its net national income. The increase in national income is partly due to an increase in tax collection and partly due to reduction in repatriation of profits by the multinational firm. The reduced repatriation also occurs from a decline in the rent of the multinational firm's capital. Of course, the positive income-augmenting effect of the tax on multinational investment should not be overdone. A large enough increase in the tax rate can drive the entire multinational capital out of the country, and the least amount of loss in real income would then be the tax revenue that would have been earned by a small tax. This suggests that there may exist an optimal t^* which is given by $dI^*/dt^*=0$.

---

[12] From Euler's equation, $L_x^*X_L^* + K_x^*X_K^* + S^*X_S^* = X^*$. Differentiating partially with respect to L^*, we get $X_{LL}^*L_x^* + X_{LK}^*K_x^* = -S^*X_{SL}^*$, which is negative as $X_{SL}^* > 0$ (see the assumption in equation (3)).

So far we have discussed the implications of the host country's tax. If, however, the tax was imposed by the source country, the results concerning resource allocation would be reversed; so would the results concerning the real wage rate and the real rental earned by local firms in two countries. However, the rental earned by the multinationals will still decline.

What is the impact of the corporate income tax in either country on the source country's national income which is defined as

$$(26)\quad I=pX+Y+(1-t^*)(p^*X^*-w^*L_x^*)$$

Expanding and substituting as in equation (24) and differentiating, we obtain

$$(27)\quad \frac{dI}{dt^*}=\frac{t}{1-t}r_x\underset{(+)}{\frac{dK_x}{dt^*}}+K_x\underset{(-)}{\frac{dr_x}{dt^*}}$$
$$-\Big(\underset{(+)}{p^*X^*-p^*X_L^*L^*-p^*X_K^*K_x^*}\Big)$$
$$-(1-t^*)\Bigg[p^*\Big(\underset{(-)}{X_{LL}^*L_x^*+X_{LK}^*K^*}\Big)\underset{(-)}{\frac{dL_x^*}{dt^*}}$$
$$+p^*\Big(\underset{(-)}{X_{KK}^*K_x^*+X_{LK}^*L_x^*}\Big)\underset{(-)}{\frac{dK_x^*}{dt^*}}\Bigg]$$

If initially there is no corporate income taxation in the source country so that $t=0$, then dI/dt^* is unambiguously negative. Differentiating (26) with respect to t,

$$(28)\quad \frac{dI}{dt}=\frac{t}{1-t}r_x\underset{(-)}{\frac{dK_x}{dt}}-(1-t^*)L_x\underset{(+)}{\frac{dw^*}{dt}}$$

showing that the national income of the source country declines unambiguously.

The results derived in this section emphatically point out the strong degree of economic interdependence that the emergence of multinational firms has brought into existence. In traditional Heckscher-Ohlin trade models, a tax on the international investment in a small-country world simply lowers the earnings on such investments while leaving almost everything else unscathed. For example, there is no change in the distribution of income within each country because such distribution is related only to relative product prices which are constant for small countries. With regard to resource allocation, the outcome is of perfunctory interest for at least that country where complete specialization is assumed—not by choice but by considerations of logical consistency or determinacy.[13] Our analysis, however, shows that the international investment tax policy of one country is bound to affect resource allocation and income distribution in both countries even in a small-country world.

A word may be said about the perennial trade union clamor in the United States for granting lower tax credits on overseas profits of the multinational firms. The workers in either country benefit from an increase in the corporate tax in the other country which will drive the multinational capital away and so increase the real wage rate in the country which did not increase the tax. Although laborers in one country do not have the political clout to demand an increase in the tax rate in another country, a reduction in the tax credit given by the source country would be equivalent to an increase in the host country's taxes (as far as multinationals are concerned), thereby benefiting the workers in the source country. The hostility of *U.S.* trade unions to the favorable tax treatment received by multinationals becomes intelligible in the light of our model.

---

[13] This problem arises in the Heckscher-Ohlin model, because factor rewards are related to the commodity price ratio, provided there is incomplete specialization. Factor supplies come into play only when there is complete specialization. For further details on this point, see Jones (1967). This problem, however, disappears in the specific factor model, because there the number of factors exceed the number of goods. See Batra and Casas, for instance.

Finally, our analysis suggests that taxes on investment by the multinational firms can be used as a means of granting protection to import-competing industries. Thus to the traditional list of devices of protectionism, such as tariffs, production subsidies and quotas, our analysis adds another tool, namely the international investment tax, that enables a country to encourage production of import-competing industries, provided, of course, that such industries are operated by the local firms.

III. Income Taxes on Local Firms

The implications of income taxes on the local firms are in sharp contrast to those of international investment taxes analyzed above: There are none. Such taxes simply lower the net rate of return earned by the local firms without causing any change in outputs or factor rewards. Since capital is specific to industries operated by the local corporations, it cannot move elsewhere in response to any taxes; similarly labor employed in local industries does not move to other industries. All that happens is that the local firms are poorer in proportion to the tax.[14]

IV. Implications of Tariffs

As with international investment taxes, the effects of tariffs turn out to be substantially different from what is known from conventional wisdom. Here again the interdependent nature of the world economies comes to light. Suppose that goods produced by the multinationals in the host country are importable goods there and that the host country introduces a tariff on its imports, so that $p^* = p(1+\tau^*)$ where τ^* is the tariff rate. Differentiating equilibrium condition (17) with respect to the tariff rate, we get

$$[A][B] = [G] \tag{29}$$

where $[A]$ and $[B]$ are the matrices in (18′) and

$$[G] = -\begin{bmatrix} X_L^* d\tau^* \\ 0 \\ X_K^* d\tau^* \end{bmatrix} \tag{30}$$

From (29) we get (assuming for convenience that $t = t^* = 0$),

$$\begin{aligned} \frac{dL_x^*}{d\tau^*} = \{ & X_L^*[p^2 H_x + pX_{KK}Y_{LL} \\ & + p^* X_{KK}^*(pX_{LL} + Y_{LL})] \\ & - X_K^*[X_{LK}^*(pX_{LL} + Y_{LL})]\} \div D \end{aligned} \tag{31}$$

$$\begin{aligned} \frac{dK_x^*}{d\tau^*} = \{ & (p^* X_{LL}^* + Y_{LL}^*) \\ & \times (pX_{LL} + Y_{LL})X_K^* - p^{*2} X_{LK}^* X_L^* \\ & \times (pX_{LL} + Y_{LL})\} \div D \end{aligned} \tag{32}$$

$$\begin{aligned} \frac{dL_x}{d\tau^*} = \{ & (p^* X_{LL}^* + Y_{LL}^*) p X_{LK} X_K^* \\ & - pp^* X_L^* X_{LK} X_{LK}^*\} \div D \end{aligned} \tag{33}$$

Here $dL_x^*/d\tau^* > 0$, $dK_x^*/d\tau^* > 0$ and $dL_x/d\tau^* < 0$. In other words, the host country's tariff induces capital to move there, employment of labor rises in X^* and so does its output. In the source country the opposite happens: Capital moves away and employment of labor in X declines and so does its output. At the same time Y^* declines and Y rises, where Y is the importable good sector in the source country. Thus, we reach an unusual conclusion: *The tariff by*

---

[14] Thus the effects of taxation on multinational and local firms are dramatically different. This asymmetry is not present in the Heckscher-Ohlin model.

one country protects the import-competing industries in both countries. This also implies that tariffs by one country are much more restrictive of international trade than has been noted before. These results make an interesting comparison with those available from the conventional models of international investment where the tariff by a small country protects its own industry while leaving output levels unchanged in the other.

Let us now proceed to the implications of the tariff for real factor rewards. With $w^* = Y^*_L$

$$\frac{dw^*}{d\tau^*} = -Y^*_{LL}\frac{dL^*_x}{d\tau^*} > 0$$

$$\frac{dw}{d\tau^*} = -Y_{LL}\frac{dL_x}{d\tau^*} < 0$$

The wage rate rises in the host country and declines in the source country. Similarly, from $r_x = pX_K$ and $K_x + K^*_x = \bar{K}_x$,

$$\frac{dr_x}{d\tau^*} = p\left[X_{KL}\frac{dL_x}{d\tau^*} + X_{KK}\frac{dK_x}{d\tau^*}\right]$$

$$= p\left[\underset{(+)}{X_{KL}}\ \underset{(-)}{\frac{dL_x}{d\tau^*}} - \underset{(-)}{X_{KK}}\ \underset{(+)}{\frac{dK^*_x}{d\tau^*}}\right]$$

Thus, a priori nothing can be said about the response of the rental on capital owned by the multinational firms. As regards the return earned by local industries, since $r_y = Y_K$ and $r^*_y = Y^*_K$, we get

$$\frac{dr_y}{d\tau^*} = -Y_{KL}\frac{dL_x}{d\tau^*} > 0$$

and $$\frac{dr^*_y}{d\tau^*} = -Y^*_{KL}\frac{dL^*_x}{d\tau^*} < 0$$

An intuitive explanation of these results is that a multinational firm shifts capital to the host country when the latter imposes a tariff. This leads to a shift of workers from Y^* to X^*. The withdrawal of labor from local industries in the host country leads to an increase in the capital-labor ratio and a fall in the rent of capital in that sector. As for the multinational firm, employment of capital and labor both increases in the host country and decreases in the source country. Because of the parallel movement of the two factors, the change in its rental is indeterminate.

In short, the implications of tariffs for income distribution in our model are completely different from the standard Heckscher-Ohlin model. There the distribution of income is determined by the Stolper-Samuelson theorem which lays emphasis on the factor-intensity differential in two industries within a country; and the tariff by one country, of course, makes no difference to factor rewards in the other country even when international investment is present.

V. Tariffs on Goods Produced by Local Firms

If the host country imports Y^* and imposes a tariff, then all the results derived above are reversed. This result contrasts with the similar case, examined in Section III, of the imposition of a tax imposed on the income earned by local firms. There, it may be recalled, the tax simply succeeded in collecting revenue from the local firms without affecting other economic variables.

In order to explore the consequences of the host country's tariff on Y^*, simply replace the first equation of system (17) by

$$p^*X^*_L = (1+\tau^*_y)Y^*_L$$

with $t = t^* = 0$, and then differentiate totally.

It is interesting to note that protection granted to local firms will drive the multinational capital out, and although r^*_y will rise, the real wage rate will decline in the host country. Thus, labor unions' proclivity for protection may turn out to be self-defeating if it is granted to goods produced by local firms.

VI. Implications of Terms of Trade

Suppose there is an exogenous change in the terms of trade. The implications of this change in the conventional model are quite simple: The output of the relatively more expensive good rises, and so does the real reward of its intensive factor, whereas the real reward of its unintensive factor declines; and all this would occur in both countries. In addition, the effects of tariffs on resource allocation and income distribution are identical to those of changes in relative product prices.

In our model incorporating the phenomenon of the multinational corporation, the results turn out to be dramatically different. Let us assume, without loss of generality, that $t=t^*=\tau^*=0$ so that $p=p^*$. Differentiating the system (17) totally with respect to p, we obtain

$$(34)\quad \frac{dL_x^*}{dp}=\{X_L^*[p^2H_x+p^2X_{LL}X_{KK}$$

$$+pX_{KK}^*(pX_{LL}+Y_{LL})]-p^2X_LX_{KL}X_{KL}^*\}\div D$$

$$(35)\quad \frac{dK_x^*}{dp}=\{pX_LX_{KL}(pX_{LL}^*+Y_{LL}^*)$$

$$-pX_L^*X_{KL}^*(pX_{LL}+Y_{LL})\}\div D$$

$$(36)\quad \frac{dL_x}{dp}=\{X_L[p^2H_x^*+p^2X_{LL}^*X_{KK}$$

$$+Y_{LL}^*(pX_{KK}^*+pX_{KK})-p^2X_L^*X_{KK}X_{LK}^*]\}\div D$$

All three equations are indeterminate in sign. Therefore, the effects of a change in the terms of trade on resource allocation, capital movements, and hence outputs in the two countries cannot be determined a priori. Similarly the response of factor rewards is also indeterminate.

Reasons for this ambiguity lie in the equation that determines the allocation of the multinational capital between the two countries. Looking at this equation in system (17), it may be observed that relative prices do not enter into this equation, although they do in the equations concerning labor allocation. The movement of capital from one country to the other is then determined by how in each country the price-induced change in labor allocation affects the marginal product of capital employed in the goods produced by the multinational firms. Suppose, for the moment, that L_x and L_x^* both rise in response to the rise in the marginal value product of labor in X and X^* generated by the rise in p. With X_{KL} and X_{KL}^* positive, the marginal product of capital rises in both X and X^*, and it is not clear which country exercises the greater pull for the multinational capital.

Indeed if the factors affecting dK_x^* cancel out, so that $dK_x^*=0$, then dL_x and dL_x^* become determinate. Because then from (35)

$$(pX_{LL}+Y_{LL})=\frac{X_LX_{KL}(pX_{LL}^*+Y_{LL}^*)}{X_{LK}^*X_L}$$

and using this we obtain

$$\frac{dL_x^*}{dp}=\left\{X_L^*(p^2H_x+pX_{KK}Y_{LL})\right.$$

$$\left.+\frac{X_LX_{KL}}{X_{KL}^*}[p^2H_x^*+pX_{KK}^*Y_{LL}^*]\right\}\div D$$

Thus dL_x^*/dp and similarly dL_x/dp are seen to be positive. If the multinational capital does not move in response to the change in the terms of trade, L_x and L_x^* go up and so do X and X^*; Y and Y^* on the contrary decline and we obtain the expected results regarding the price output response.

The same restriction applies if we are interested in obtaining a clear-cut response of factor rewards to the change in p. If dL_x^*/dp and dL_x/dp are both positive, then the wage rate rises in both countries, because $dw^*/dp=-Y_{LL}^*(dL_x^*/dp)$ and $dw/dp=-Y_{LL}(dL_x/dp)$ are both positive. Similarly r_x rises but r_y and r_y^* decline, because

$$\frac{dr_x}{dp}=X_K+X_{KL}\frac{dL_x}{dp}$$

$$\frac{dr_y}{dp}=-Y_{KL}\frac{dL_x}{dp}$$

$$\frac{dr_y^*}{dp}=-Y_{KL}^*\frac{dL_x^*}{dp}$$

All these results confirm what has been stated before, that the implications of tariffs on the one hand and those of the terms of trade on the other are dramatically different from each other. This is a novelty that flows from our model that incorporates some features of the multinational industries.

VII. Conclusions

In this paper we have developed a general equilibrium model that incorporates many features of an international economy where the multinational firms have generated economic conditions which deserve a careful analysis. The framework developed here explains many observed phenomena that the conventional models of international trade and investment are not equipped to tackle. The model stresses the extreme degree of economic interdependence that the institution of multinational enterprise has ushered. All our results corroborate this central theme: that international economic policies pursued by one nation make waves in other nations even in a small-country world where relative prices are determined in world markets, with each nation taking the terms of trade as constant. Some of our results confirming this view are summarized below.

1. An income tax (or profit tax) on the multinational firms by any country drives their capital out and lowers their rate of return on capital stock and possibly on their specific factors, such as technology, managerial know-how, marketing techniques, etc.; the real wage rate declines in the tax-imposing country but rises in the other country. The international investment tax (or subsidy) can also be used for granting protection to import-competing industries —a policy device that requires no intervention in the goods market.
2. The profits tax on the local firms has no effect on other variables in any country; it merely lowers the rate of return on capital earned by the taxed sector.
3. One country's tariff stimulates the output of import-competing industries of not only that country but also of the other country. The real wage rate rises if the tariff is imposed on goods produced by the multinational firms but declines if the local firms are the beneficiaries of protection. The opposite occurs in the other country.
4. The implications of an exogenous change in relative prices are different from a similar change effected by the tariff. This is an unusual feature of our model and is not present in other conventional models of international trade.

REFERENCES

Robert E. Baldwin and J. David Richardson, *International Trade and Finance*, Boston 1974.

Richard J. Barnett and Ronald E. Muller, *Global Reach: The Power of the Multinational Corporations*, New York 1974.

R. N. Batra and F. R. Casas, "A Synthesis of Heckscher-Ohlin and Neoclassical Models of International Trade," *J. Int. Econ.*, Mar. 1976, *6*, 21–38.

______ **and J. Hadar,** "Theory of the Multinational Firm: Fixed Versus Floating Exchange Rates," *Oxford Econ. Pap.*, July 1979, *31*, 258–69.

R. E. Caves, "International Corporations: The Industrial Economics of Foreign Investment," *Economica*, Feb. 1971, *38*, 1–27.

T. Horst, "The Theory of the Multinational Firm: Optimal Behaviour under Different Tariff and Tax Rates," *J. Polit. Econ.*, Aug. 1971, *79*, 1059–72.

H. G. Johnson, "The Efficiency and Welfare Implications of the International Corporation," in Charles P. Kindleberger, ed., *The International Corporation*, Cambridge, Mass. 1970.

R. W. Jones, "The International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–38.

______, "A Three Factor Model in Theory, Trade and History," in Jagdish Bhagwati, et al., eds., *Trade, Balance of Payments and Growth: Essays in Honor of C. P. Kindleberger*, Amsterdam 1971.

Murray C. Kemp, *Theory of International Trade and Investment*, Englewood Cliffs 1969.

H. Leibenstein, "Allocative Efficiency vs. X-Efficiency," *Amer. Econ. Rev.*, June 1966, *56*, 392–415.

______, "X-Inefficiency Xists—Reply to an Xorcist," *Amer. Econ. Rev.*, Mar. 1978,

P17822

68, 203–11.

C. F. Pratten, *Labour Productivity Differentials within International Companies*, Cambridge 1976.

Sidney E. Rolfe, "The International Corporation," Background Report, XII Congress, International Chamber of Commerce, Istanbul 1969.

W. F. Stolper and P. A. Samuelson, "Protection and Real Wages," *Rev. Econ. Stud.*, Nov. 1941, *9*, 58–73.

Jean J. Servan-Schreiber, *The American Challenge*, New York 1968.

T. Y. Shen, "Technology Diffusion, Substitution, and X-Efficiency," *Econometrica*, Mar. 1973, *41*, 263–84.

Jacob Viner, *The Customs Union Issue*, New York 1950.

The Terms of Trade and Equilibrium Growth in the World Economy

By RONALD FINDLAY*

The extensive discussions that have been going on in connection with proposals for a New International Economic Order have focused attention on the nature of the economic relations between the advanced and less developed regions of the world economy, or the "North" and "South" as it has become customary to refer to them. Ever since the work of Raul Prebisch and Hans Singer almost three decades ago the movement of the terms of trade between these regions has been regarded as a key index of the distribution of the benefits from the international division of labor and the development prospects of the South. Although an extensive and controversial literature has developed around these issues, there has been a significant lack of rigorous formal analysis to provide a framework within which the diverse arguments and pieces of empirical evidence can be sorted out and assessed.

The objective of this paper is to present a highly stylized dynamic model in which the terms of trade emerge as the mechanism linking the growth rates of output in the North and South. The analysis attempts to forge a synthesis of three famous papers, using the neoclassical growth model of Robert Solow for the North, the "labor surplus" or "dual economy" model of W. Arthur Lewis for the South, and the Harry Johnson (Part 2) treatment of the terms of trade to serve as the link between them.

I

Imagine a world economy consisting of two regions, the North and the South. The North produces a single composite commodity, manufactures, which can be used for either consumption or investment. The technology for manufactures is represented by a neoclassical production function with constant returns to scale and with capital and labor as the inputs. Capital consists of a stock of manufactures. The size of the labor force is given at any instant and grows over time at a constant rate. There is also labor-augmenting technological progress at a fixed rate assumed to be going on. Under these assumptions, as is well known, the production function can be expressed in terms of capital and "effective" labor, which grows over time at a fixed rate equal to the sum of the rate of natural increase and the rate of technological progress. A constant fraction of output is saved and invested. Markets are perfectly competitive and are cleared instantaneously so that there is always full employment of labor and full utilization of capacity.

These assumptions, of course, correspond exactly to the Solow neoclassical growth model. The only change is that the portion of output in the North which is not saved is spent either on manufactures or on another homogeneous commodity, representing primary products, with the proportions depending upon the relative price of the two goods. Income elasticities of demand for both goods are unity.

Primary products constitute the sole output of the South. There is again a neoclassical production function with capital and labor as the inputs governing the output of primary products. Capital consists of a stock of manufactures, just as in the North. Labor is in perfectly elastic supply from a "hinterland," which is otherwise outside the model, at a fixed real wage in terms of primary products. There is perfect competition so labor is hired up to the point at which its marginal productivity is equal to the fixed real wage. Total employment at any instant will depend upon the quantity of capital

---

*Columbia University.

available. The capital-labor ratio is uniquely determined by the fixed real wage and in turn determines the marginal product of capital and hence the amount of profit per unit of labor. The *rate* of profit, however, cannot be determined unless the relative price of manufactures and primary products is known, since capital consists of a stock of manufactures and profits of a flow of primary products.

It is assumed that a constant fraction of profits are saved and all wages are consumed. Consumption expenditure, whether out of wages or profits, is spent on manufactures and primary products with the proportions depending upon relative prices. Income elasticities of demand for both goods are unity. Investment demand is thus purely for manufactures while consumption demand is divided between the two goods. The fraction of profits which is saved determines the increase in the stock of capital that propels the system forward.

It is apparent that this depiction of the South corresponds closely to the celebrated Lewis model of economic development with unlimited supplies of labor. The Solow economy of the North and the Lewis economy of the South are linked through international trade. It is assumed that there is no lending so that trade is always balanced. The demand for imports in each region depends upon the relative price of the two goods and real income. Johnson obtained a simple formula for the change in the terms of trade resulting from exogenously given growth rates of output in the two regions. In the present model, however, there is a feedback from the terms of trade to the growth rate of output in the South. The next section will demonstrate how the terms of trade and the growth rate of the South have to adjust to the growth rate of the "effective" labor force in the North that determines the steady-state growth of the whole interdependent dynamic system that constitutes the specification of the world economy.

II

The basic structure of the model will be presented in this section. The following notation will be used:

q = output per worker of manufactures in the North
π = output per worker of primary products in the South
L_N, L_S = employment in North and South
Y_N, Y_S = total incomes in North and South
k_N, k_S = capital-labor ratios in North and South
$\overline{w}$ = fixed wage in the South
n = fixed growth rate of effective labor in the North
θ = the terms of trade (manufactures per unit of primary products)
λ = ratio of employment in the South to employment in the North
ρ = rate of profit in the South
s = constant fraction of income saved in the North
σ = constant fraction of profit saved in the South
I_N, I_S = total imports in North and South
m = per capita imports of the North
μ = per capita consumption imports of the South
η_N, η_S = elasticities of m and μ with respect to θ and $1/\theta$, respectively.

Asterisks will denote long-run equilibrium values of the corresponding variables.

The constant returns-to-scale production function for primary products in the South is

$$\pi = \pi(k_S) \tag{1}$$

and the profit-maximization condition is

$$\pi(k_S) - \pi'(k_S)k_S = \overline{w} \tag{2}$$

Under the usual assumptions that $\pi'(k_S) > 0$ and $\pi''(k_S) < 0$, there is a unique k_S^* that satisfies (2). Since capital consists of a stock of manufactures and output of a flow of primary products, the rate of profit in the South is

$$\rho = \theta\pi'(k_S^*) \tag{3}$$

while the common growth rate of total capital, output, and employment in the South is

$$g = \sigma\rho \tag{4}$$

which is the so-called "Anglo-Italian" equation.

A necessary condition for a steady-state equilibrium in the world economy is that the growth rates of North and South should be equal. Since the natural growth rate of the North is the constant n, it follows from (3) and (4) that there is a unique value

$$\theta^* = \frac{n}{\sigma\pi'(k_S^*)} \tag{5}$$

of the terms of trade that is consistent with balanced growth of the world economy as a whole. Notice that θ^* depends only on $n, \sigma, \bar{w}$, and the production function for the South. It is completely independent of the production function and propensity to save of the North, and of the demand for imports in the two regions.

The terms of trade must also, however, equilibrate the balance of trade. I now turn to the determination of the terms of trade in the "short run" by reciprocal supply and demand for exports and imports in the usual manner, after which it will be shown how the short- and long-run values of the terms of trade become reconciled with each other.

The production function for the North is

$$Y_N = q(k_N)L_N \tag{6}$$

where L_N depends upon time by the relation

$$L_N = L_0 e^{nt} \tag{7}$$

and k_N at any instant is determined by past history. The import demand function of the North is

$$I_N = m[\theta, (1-s)q(k_N)]L_N \tag{8}$$

where it is assumed that $m'(\theta) < 0$ and that the elasticity of m with respect to per capita consumption $(1-s)q(k_N)$ is equal to unity. The import demand function for the South is

$$I_S = \left[\theta\sigma\pi'(k_S^*)k_S^* + \mu\left(\frac{1}{\theta}, \bar{w} + (1-\sigma)\pi'(k_S^*)k_S^*\right)\right]L_S \tag{9}$$

where the expression in brackets is the sum of per capita imports for investment and consumption, respectively. At any price $1/\theta$ the demand for capital goods will be $\theta\sigma\pi'(k_S^*)k_S^*$ per worker employed, so that investment is always equal to the fraction σ of profits in terms of primary products, implying that the price elasticity of demand for capital goods imports is unity. Per capita import demand for consumption purposes is μ, with $\mu'(1/\theta) < 0$ and the elasticity with respect to per capita consumption $\bar{w} + (1-\sigma)\pi'(k_S^*)k_S^*$ is equal to unity. Southern employment L_S is determined by

$$k_S^* L_S = K_S \tag{10}$$

where the total capital stock at any instant K_S is given by past history. In the absence of international capital movements the trade balance must be zero so that

$$\theta I_N = I_S \tag{11}$$

Substituting for I_N and I_S from (8) and (9) and letting $\lambda = L_S/L_N$ denote the ratio of employment in the two regions, (11) can be solved to obtain

$$\theta = \frac{\lambda\mu\left[\left(\frac{1}{\theta}\right), \bar{w} + (1-\sigma)\pi'(k_S^*)k_S^*\right]}{m(\theta, k_N) - \lambda\sigma\pi'(k_S^*)k_S^*} \tag{12}$$

In the short run λ and k_N are given by past history so that (12) determines the value of θ that equilibrates the trade balance. On the other hand, if θ is at its long-run value θ^* as determined by (5) then for each k_N there is a value of λ such that (12) is satisfied. In particular, if k_N is at the steady-state value k_N^* given by the condition

$$k_N^* = \frac{sq(k_N^*)}{n} \tag{13}$$

then (12) determines the corresponding unique steady-state value of λ as

$$\lambda^* = \frac{\theta^* m^*(\theta^*, k_N^*)}{\left[\mu^*\left(\frac{1}{\theta^*}, \bar{w} + (1-\sigma)\pi'(k_S^*)k_S^*\right) + nk_S^*\right]} \tag{14}$$

after using (5). It is apparent that λ^* is equal to the ratio of steady-state per capita imports for the two regions.

The ratio of total incomes in the two regions is

$$(15) \qquad \frac{Y_S}{Y_N} = \frac{\theta^* \pi(k^*)\lambda^*}{q(k_N^*)}$$

In interpreting (15) it should be noted that $q(k_N^*)$ is output per unit of "effective labor." If n includes labor-augmenting technical progress then the ratio of effective to actual labor will be continually increasing in the North. The per capita income of the actual labor force will be continually rising in the North while it will be constant in the South since n is there the rate of growth of the actual labor force.

III

Having obtained the steady-state solution of the model, this section examines the question of convergence towards the steady state from arbitrary initial conditions. At any instant L_N, K_N, and K_S are all given by past history, and L_S is determined by the fixed wage $\bar{w}$ and the corresponding capital-labor ratio k_S^*. Dividing K_N and L_S by L_N, one obtains the normalized variables k_N and λ which are the state variables of the dynamic system to be studied explicitly in this section. Capital in the South K_S is not a separate state variable since it is always strictly proportional to L_S by virtue of the fixed-wage assumption. Under the assumptions about saving out of profits and wages, the common growth rate of K_S and L_S is $\sigma\pi'(k_S^*)\theta(\lambda, k_N)$. The relative rate of change of λ is therefore the difference between this rate and the fixed growth rate n of labor force in the North. The rate of change of k_N is equal to the difference between per capita saving and the requirements for maintaining the capital per head of the growing labor force.

The basic dynamic system can therefore be formulated as

$$(16) \qquad D\lambda = [\sigma\pi'(k_S^*)\theta(\lambda, k_N) - n]\lambda$$

$$(17) \qquad Dk_N = sq(k_N) - nk_N$$

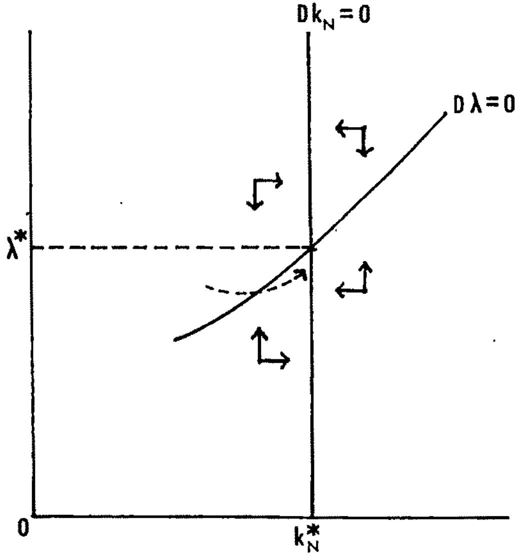


FIGURE 1

where D denotes the derivative with respect to time.

The properties of this system can be conveniently seen by means of the phase diagram constructed in Figure 1. The $Dk_N = 0$ locus is a vertical line at the value k_N^* obtained in (13). It is independent of λ. The condition for $D\lambda = 0$ is that θ should be at the value θ^* obtained earlier in (5). The $D\lambda = 0$ locus is therefore one which shows the value of λ corresponding to any value of k_N such that $\theta = \theta^*$. The positive slope of the $D\lambda = 0$ locus is derived from the relation

$$(18) \qquad \left.\frac{d\lambda}{dk_N}\right|_{\theta=\theta^*} = -\frac{\partial\theta}{\partial k_N} \Big/ \frac{\partial\theta}{\partial\lambda}$$

which is obtained by total differentiation of (12) evaluated at θ^*, with

(19)

$$\frac{\partial\theta}{\partial k_N} = -\frac{(1-s)q'(k_N)\tilde{m}\theta}{[m - \lambda\sigma\pi'(k_S^*)k_S^*]\left(1 - \eta_S - \frac{\theta m}{\lambda\mu}\eta_N\right)}$$

(20)

$$\frac{\partial\theta}{\partial\lambda} = \frac{[\mu + \theta\sigma\pi'(k_S^*)k_S^*]}{[m - \lambda\sigma\pi'(k_S^*)k_S^*]\left(1 - \eta_S - \frac{\theta m}{\lambda\mu}\eta_N\right)}$$

The common denominator of (19) and (20) is the product of two terms, the first of which is the denominator of (12) and is positive. The second is negative if

$$\eta_S + \frac{\theta m}{\lambda\mu}\eta_N > 1 \tag{21}$$

where η_S and η_N are the elasticities of consumer import demand with respect to the terms of trade in the two regions. This is what corresponds to the familiar Marshall-Lerner stability condition in the present model, since the elasticity of demand for investment imports in the South is unity. It is apparent that (21) and the Marshall-Lerner condition are equivalent since (21) implies that η_N plus a weighted average of η_S and the unit elasticity of investment imports with $\lambda\mu/\theta m$ as the weight of η_S exceeds unity. The symbol $\tilde{m}$ in the numerator of (19) is the partial derivative of m with respect to per capita consumption expenditure $(1-s)q(k_N)$ in (8). It follows that

$$\frac{\partial\theta}{\partial k_N} > 0, \quad \frac{\partial\theta}{\partial\lambda} < 0$$

if (21) holds, which establishes the positive slope of the $D\lambda = 0$ locus from (18). The reason for this slope is clear since an increase in k_N improves the terms of trade of the South above θ^* so that an increase in λ is required to restore this value. The intersection of the $D\lambda = 0$ and $Dk_N = 0$ loci determines the steady-state value λ^* of λ, already obtained in (14).

Examining the dynamic system (16) and (17) the "trace condition" for stability is seen to be satisfied since

$$\frac{\partial(D\lambda)}{\partial\lambda} = \lambda^*\sigma\pi'(k_S^*)\frac{\partial\theta}{\partial\lambda} < 0 \tag{22}$$

$$\frac{\partial(Dk_N)}{\partial k_N} = sq'(k_N^*) - n < 0 \tag{23}$$

at the steady-state point (λ^*, k_N^*). The condition (23) is the same as the familiar stability condition for the Solow model of a closed economy.

The "determinant condition" for stability is also fulfilled since

$$\frac{\partial(D\lambda)}{\partial k_N} = \lambda^*\sigma\pi'(k_S^*)\frac{\partial\theta}{\partial k_N} > 0 \tag{24}$$

$$\frac{\partial(Dk_N)}{\partial\lambda} = 0 \tag{25}$$

The conditions imply that λ falls at any point above the $D\lambda = 0$ locus and rises at any point below it, while k_N increases at any point to the left of the $Dk_N = 0$ locus and decreases at any point to the right. The motion of λ and k_N is indicated by the arrows in each of the quadrants. A possible trajectory from an arbitrary initial condition is indicated by the dashed curve in Figure 1. At the initial point θ is below θ^* so that the growth rate of the South is below n which means that λ falls and θ increases. When the $D\lambda = 0$ locus is reached $\theta = \theta^*$. The increase in k_N raises the terms of trade above θ^* which means that λ must now start to increase together with k_N towards the steady-state values λ^* and k_N^* while θ approaches θ^* from above. The terms of trade therefore overshoot the long-run value θ^* but then are pulled towards it. The intersection of the trajectory with the $D\lambda = 0$ locus divides the approach towards the steady state into two phases, the first in which the South grows more slowly than the North and the terms of trade continually improve, and the second in which the South grows faster than the North and the terms of trade converge toward θ^* from above.

IV

This section investigates the impact and long-run effects of variations in the parameters. The initial position will in all cases be taken as the steady-state point (λ^*, k_N^*). First, suppose that there is a shift of demand in the North towards primary products, so that m is larger for any given θ. The impact effect on θ is given by partial differentiation of (12) with respect to a shift parameter which reveals that θ would increase if the stability condition (21) holds. This raises the rate of profit in the South which causes λ to increase and θ to fall. It is clear from (5) that the long-run value of θ^* is independent of the import demand function of the North, so that θ converges back towards the original θ^*, while (14) indicates that λ^* increases in the same proportion as the increase in m. In terms of Figure 1 the effect is to shift the $D\lambda=0$ schedule upwards, while leaving the $Dk_N=0$ schedule unchanged, so that λ^* rises and k_N^* remains constant.

The impact and long-run effects of the shift in import demand can be conveniently seen in terms of Figure 2. The BB curve shows the combinations of θ and λ that maintain equilibrium in the trade balance when the capital-labor ratio of the North is at its steady-state value k_N^*. The negative slope follows from the assumption that the Marshall-Lerner stability condition (21) holds. An increase in the propensity to import of the North shifts BB to the right to $B'B'$, since λ has to increase at constant θ for (12) to still hold. In the short run λ cannot change so it remains at the original value of λ^* while θ increases to the corresponding point on $B'B'$. This raises the rate of growth of the South so λ increases gradually while θ falls and the economy of the South moves down $B'B'$, approaching the point on it corresponding to the original terms of trade θ^*. The same line of reasoning indicates that a shift in the South's consumption demand towards manufactures eventually leaves θ^* unchanged but reduces λ^* by a proportion equal to the product of the increase in consumption demand and the share of imports for consumption to total imports in the South. The effects on relative income shares of these shifts in import demands are proportional to the changes in λ^*, as (15) indicates.

An increase in the propensity to save of the North raises k_N^* by virtue of (13) so that the $Dk_N=0$ schedule is shifted to the right, while (5) shows that θ^* is not affected by this change. An increase in s, for given values of θ and k_N, must lead by (12) to a reduction in λ, so that the $D\lambda=0$ schedule is shifted downwards as a result. This is because a rise in s means a shift towards manufactures and away from primary products in the demand pattern of the North, so that λ must fall if there is to be equilibrium in the balance of trade with everything else held constant. The fact that the $D\lambda=0$ locus shifts down while the $Dk_N=0$ locus shifts to the right might appear to render the effect of a rise in s on λ^* ambiguous. If the North is to the left of the Golden Rule point where $q'(k_N)=n$, however, a rise in s must raise total consumption per capita $(1-s)q(k_N^*)$. It follows from (14) that λ^* must increase in the same proportion as steady-state per capita consumption in the North, since the elasticity of demand for primary products with respect to consumption expenditure is unity. It is interesting to note, therefore, that λ^* is maximized, for given values of all other parameters, when the propensity to save in the North is at the level required by the Golden Rule. The disparity in per capita consumption and income levels between North and South is of course widened by the increase in s since the levels of these variables in the South are not affected by s. The ratio of *total* incomes shown in (15) worsens for the South in spite of the increase in λ^*, since the proportionate increase of this variable is equal to $(1-s)q(k_N^*)$ which is less than that of $q(k_N^*)$ because of the rise in s. An increase in the North's propensity to save therefore raises relative employment in the South but

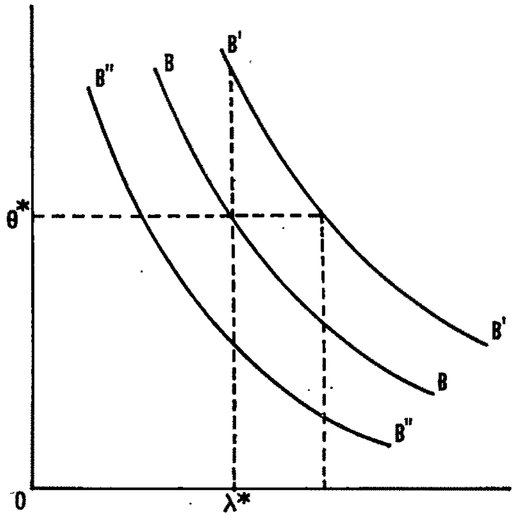


FIGURE 2

it lowers its relative per capita income. The favorable effect on relative total income of the South of the improvement in relative employment is not sufficient to offset the decline in relative per capita incomes.

Technological change in the North can be introduced in the form of a once and for all Hicks-neutral shift in the production function $q(k_N)$, or in the form of an increase in the rate of labor-augmenting Harrod-neutral progress that results in an increase in n, the growth rate of the effective labor force. Taking the case of the Hicks-neutral shift first, it is evident that the $Dk_N=0$ locus is shifted to the right and that k_N^* increases. For given values of θ and k_N (12) implies that λ must increase if $q(k_N)$ rises. The $D\lambda=0$ locus is therefore shifted upward and so λ^* increases. The immediate impact of the Hicks-neutral shift in the production function of the North is for θ to increase with λ^* and k_N^* at their original steady-state values. The initial rise of θ raises the growth rate of the South above n, which causes λ to increase towards its new steady-state value, while k_N also increases because savings in the North now exceed the steady-state requirement. The fact that the South grows faster than the North means that θ falls after the initial jump in its level, and it eventually approaches its original level θ^* which, as may be seen from (5), is not affected by the shift in $q(k_N)$. It can be seen from (14) and (15) that the shift in $q(k_N)$ has no effect on the relative total incomes of the North and South since λ^* increases in the same proportion as $q(k_N)$ while θ^* remains unchanged. Relative per capita incomes of course move in favor of the North.

An increase in n shifts the $Dk_N=0$ locus to the left and reduces k_N^*. From (5) it is clear that θ^* must rise in the same proportion as n, so that there is a permanent improvement in terms of trade of the South. For any given k_N it follows from (12) that a higher θ^* requires a lower λ to equilibrate the trade balance, so that the $D\lambda=0$ schedule is shifted downward and λ^* falls.

A change in any of the South's parameters does not affect the $Dk_N=0$ locus and so leaves k_N^* unchanged. The welfare of the North, however, can be altered by these parameter shifts through their effects on the terms of trade. An increase in the fixed real wage $\bar{w}$ raises the capital-labor ratio k_S^* in the South and so reduces $\pi'(k_S^*)$, the marginal productivity of capital. This results by (5) in a rise in θ^* proportionately equal to the fall in $\pi'(k_S^*)$. The South is therefore able to improve its terms of trade in the long run by an exogenous increase in the fixed real wage. The price for this, however, will be a reduction in the relative employment ratio λ^*. This can be proved by showing that an increase in $\bar{w}$ must reduce λ^* if the terms of trade are held constant, so that the improvement in the South's terms of trade must reduce it even further by virtue of the fact that $\partial\lambda/\partial\theta<0$. The rise in k_S^* induced by the rise in $\bar{w}$, combined with the fact that the South is to the left of the Golden Rule since $\sigma<1$ means that total consumption expenditure $\bar{w}+(1-\sigma)\pi'(k_S^*)k_S^*$ and investment nk_S^* both rise on a per capita basis. At constant terms of trade this means an increase in the demand for imports for both consumption and investment purposes. Exports will however be constant, since the North's income is not affected, so that λ^* must fall as a result of the rise in $\bar{w}$ with θ^* held constant. But θ^* in fact rises, so that λ^* must fall even further since $\partial\lambda/\partial\theta<0$.

An increase in σ, the propensity to save out of profits in the South, leads by (5) to a fall in θ^* in the same proportion. The effect on λ^* is obtained by logarithmic differentiation of (14) with respect to σ, using the fact that the elasticitiy of θ^* with respect to σ is minus unity. This establishes that

$$\frac{\sigma}{\lambda^*}\frac{d\lambda^*}{d\sigma}\gtrless 0 \tag{26}$$

depending upon whether

$$\frac{\theta^* m}{\lambda^*\mu}\eta_N+\eta_S\gtrless 1+\frac{(\theta^*-\tilde{\mu})\sigma\pi'(k_S^*)k_S^*}{\mu} \tag{27}$$

where $\tilde{\mu}$ is the additional import of manufactures for consumption resulting from an increase in total consumption expenditure in terms of primary products $\bar{w}+(1-\sigma)\cdot\pi'(k_S^*)k_S^*$. The condition (27) on the sum of the elasticities of import demands is more stringent than the stability condition (21), since the marginal propensity to import when income in the South is measured in terms of manufactures is $\tilde{\mu}/\theta^*$, which is between zero and unity, making $(\alpha^* - \tilde{\mu})$ positive.

The effect of a rise in σ on λ^* at constant terms of trade is for it to be reduced, since all of the increase in saving out of profits is spent on imports while only part of what was formerly being consumed was spent on imports, requiring λ^* to fall to equilibrate the trade balance. The terms of trade worsen, however, which, if (21) holds, leads to a rise in λ^*. The necessary and sufficient condition for the induced increase in λ^* to more than offset the fall at constant terms of trade is provided by (27).

In terms of Figure 2 the increase in σ shifts the BB curve to the left to $B''B''$. In the short run λ^* remains fixed so θ falls to the corresponding point on $B''B''$. This decline, however, may be smaller or greater than the long-run decline in θ^* which must be in the same proportion as the increase in σ. The new steady-state value of λ^* will be at the point on $B''B''$ corresponding to this long-run decline in θ^* and the economy of the South moves along $B''B''$ in whichever direction is required to get to this point since $D\lambda$ will be positive or negative depending upon whether the impact effect on θ is less or more than the long-run effect.

The effect of the increase in σ is to *worsen* the relative per capita income of the South since θ^* falls while $\pi(k_S^*)$ and $q(k_N^*)$ are unchanged. The effect on relative total incomes depends in addition upon what happens to λ^*. So long as the total elasticity of λ^* with respect to σ, obtained earlier in connection with (27), does not exceed unity the relative total income of the South must also decline.

A simple way in which a once and for all technological change in the South can be introduced is by an increase in the marginal productivity of capital with the capital-labor ratio and the marginal productivity of labor unchanged. This will lower θ^* in the same proportion as the rise in the marginal productivity of capital, as implied by (5). Denoting $\pi'(k_S^*)$ by r, the effect of this type of technical progress on λ^* is obtained by logarithmic differentiation of (14) with respect to r, using the fact that the elasticity of θ^* with respect to r is minus unity. This yields the result that

$$\frac{r}{\lambda^*}\frac{d\lambda^*}{dr} \gtrless 0 \tag{28}$$

depending on whether

$$\frac{\theta^* m}{\lambda^* \mu}\eta_N + \eta_S \gtrless 1 + \frac{[\theta^* - (1-\sigma)\tilde{\mu}]rk_S^*}{\mu} \tag{29}$$

The interpretation of (29) is analogous to that of (27). The impact and long-run effects on θ and λ can also be shown in terms of Figure 2 in a manner corresponding exactly to the preceding analysis of the increase in the propensity to save, since the increase in the South's income at constant terms of trade requires λ to be reduced, thus shifting BB to the left.

The effect of this type of technological change is to *reduce* the relative per capita income of the South. The reason is that θ^* deteriorates in the same proportion as r increases while $\pi(k_S^*)$ increases only by the share of profit times this amount. Thus the relative total income of the South would also decline unless the total elasticity of λ^* with respect to r, obtained in connection with (29), were to exceed the share of wages times the proportionate increase in r resulting from the technological change.

V

"An engine of growth" was the famous phrase that Dennis Robertson used to describe the role of international trade in the expansion of the world economy from the mid-nineteenth century to World War I, a process that has continued in the last three decades after the disruptions of the interwar years. In the model that has been presented here, trade is indeed an engine of growth for the economy of the South, but the power that drives that engine is generated by the exogenously determined natural growth rate of the North. The South does not have a given natural growth rate but an endogenous one, depending upon the value of the terms of trade since this is what determines the rate of profit in the Anglo-Italian equation. The steady-state level of the terms of trade must be whatever is necessary to make the endogenous growth rate of the South equal to the fixed natural growth rate of the North.

The structural differences in the determination of the growth rate of the two regions produces asymmetrical consequences on the terms of trade of changes in technology and the propensity to save. As demonstrated in the previous section, once and for all improvements in the production function of the North and increases in its propensity to save leave the terms of trade unchanged in the long run and increase its real per capita income. In the South, on the other hand, these shifts lead to a proportionate fall in the long-run terms of trade and a decline in real per capita income measured in terms of manufactures. These unfavorable effects, however, are compensated by an increase in the relative employment ratio if the elasticities of import demand are sufficiently high. This asymmetry is a prominent feature in the writings of Raul Prebisch and Hans Singer. The arguments put forward for this asymmetry by these writers, however, usually involve the role of monopolies and trade unions in the North and generally have not been very convincing. The present analysis obtains similar conclusions in spite of assuming perfectly competitive markets everywhere.

In conclusion some of the major limitations of the model may be noted. One of these is the absence of international capital mobility in response to differences in the rate of profit between North and South. This can readily be handled but an adequate treatment requires a separate paper. Another is the assumption of complete specialization in both regions. Two-sector open dual economics have been studied, for example in my book, Part 2, though only in the context of either constant terms of trade or an exogenously shifting demand curve for the exportable primary product. Once again, an extension of the analysis to accommodate this feature is quite feasible but would require too much space to attempt here.

The most basic assumption of all, of course, has been that the South enjoys unlimited supplies of labor at a fixed real wage. This is not intended as a permanent state of affairs, since if the growth rate of employment in the "modern" or commercialized sector exceeds the rate of population growth, the unlimited supplies of labor must become exhausted and growth will eventually produce a rise in the real wage. Structural differences between North and South would then no longer exist and the neoclassical model would come into its own the world over. Hopefully that day will come, but it has not as yet.

REFERENCES

Ronald Findlay, *International Trade and Development Theory*, New York 1973.

Harry G. Johnson, *International Trade and Economic Growth*, Cambridge, Mass. 1967.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester Sch. Econ. Soc. Stud.*, May 1954, *22*, 139–91.

Raul Prebisch, *The Economic Development of Latin America and its Principal Problems*, New York 1950.

D. H. Robertson, "The Future of International Trade," *Econ. J.*, Mar. 1938, *48*, 1–14.

H. W. Singer, "The Distribution of Gains between Borrowing and Investing Countries," *Amer. Econ. Rev. Proc.*, May 1950, *40*, 473–85.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 65–94.

Taxing Tar and Nicotine

By JEFFREY E. HARRIS*

Consumer misperception of health consequences is often invoked to justify government intervention in cigarette smoking. Rather than debating this issue here, I address a more practical problem. Suppose that government intervention is warranted by such misperceptions. What then is the most appropriate corrective action?

In particular, I analyze one form of corrective action—the taxation of cigarettes according to their tar and nicotine contents. I set up a benefit-cost framework in which smokers fail to perceive the health costs of cigarette use. Within this framework, I focus on the administratively simple case where different uniform tax rates apply to cigarette brands whose tar and nicotine contents exceed or fall below a specified cutoff value. The main regulatory design question is: What should determine our choice of these two tax rates and the cutoff value of tar and nicotine? The empirical answer below strikingly illustrates how biological and economic facts interrelate in the formulation of a discriminating public policy toward cigarette smoking.

I. Assumptions

The population contains a large number of cigarette smokers, who may differ in their cigarette preferences. In making their consumption decisions, these smokers fail to take account of the health damage produced by smoking. The purpose of the government's tax policy is to correct for these failures. The magnitude of this health damage depends upon the number and type of cigarettes smoked, but it is otherwise independent of tastes and incomes.

Social welfare is defined to be the sum of smokers' willingness to pay for cigarettes minus health damages, plus cigarette producers' profits and plus government tax revenues. This benefit-cost criterion is not sensitive to the distribution of benefits among consumers, government, and cigarette producers. It is also insensitive to the distribution of health damages across smokers.

I shall index smokers' cigarette tastes by the continuous scalar parameter n, which has probability density $f(n)$. Each smoker's consumption is characterized by x, the number of cigarettes smoked; and α, an index of the tar and nicotine (*TN*) delivery of the particular brand smoked. Smokers' tastes are ordered so that in equilibrium α increases with n. That is, $\alpha_n \geqq 0$, where subscripts stand for partial derivatives. Cigarette brands, I assume, are available over a continuum of values of α.

Let $H(x,\alpha)$ denote the individual health damage, in dollars, from smoking x cigarettes with *TN* content α, where $H_x > 0$ and $H_\alpha > 0$. Let c be the marginal cost of cigarettes, which is assumed to be constant and independent of α. Let Y and y be a smoker's gross income and his income available for all other goods, respectively. Let the utility function of a smoker with preferences of type n take the form

$$(1) \qquad V(x,\alpha,y,n) = U(x,\alpha,n) + y$$

where $U(x,\alpha,n)$ is concave in x and α. Given (1), a smoker's demand for cigarettes does not depend on his gross income or on the demands for other goods. Moreover, I assume that for all n, $U_\alpha(x,\alpha,n) = 0$ for some values of x and α. That is, smokers can be satiated with respect to the *TN* delivery of cigarettes. For purely notational convenience, and without loss of generality, all smokers are assumed to have the same gross income Y.

---

*Assistant professor, Massachusetts Institute of Technology, and clinical associate, Massachusetts General Hospital. Critical comments by P. Diamond, J. Hausman, K. Roberts, D. Romer, M. Rothschild, P. Samuelson, and P. Temin are gratefully acknowledged. This study was supported by the Health Sciences Fund. The opinions expressed in this paper are my sole responsibility.

II. Uniform Price Case

First consider the case where the after-tax unit price p of cigarettes is independent of α. This case closely approximates current cigarette pricing and taxation.

Each smoker, failing to perceive the cost H, chooses (x,α,y) to maximize V subject to the budget constraint $px+y=Y$. From (1), this is equivalent to choosing (x,α) to maximize $U(x,\alpha,n)-px+Y$. A smoker's contribution to government tax revenues and cigarette producers' profits is $(p-c)x$. Social welfare is therefore

$$(2)\quad W=\int_R [V(x,\alpha,y,n)+(p-c)x - H(x,\alpha)]f(n)dn$$
$$=\int_R [U(x,\alpha,n)+Y-cx-H(x,\alpha)]f(n)dn$$

where the domain of integration is the real line R, and where x, α, and y depend upon p and n.

For each cigarette smoker of type n, the necessary first-order conditions for utility maximization include

$$(3)\qquad U_x(x,\alpha,n)=p$$
$$U_\alpha(x,\alpha,n)=0$$

Denote the utility-maximizing choices solving (3) by $x(p,n)$ and $\alpha(p,n)$. Although $x_p(p,n)\leqq 0$, all p,n, the sign of $\alpha_p(p,n)$ may be positive or negative. When $\alpha_p<0$, cigarette TN and smoking frequency are complements, and when $\alpha_p>0$, they are substitutes. Differentiate W with respect to p and substitute (3)

$$(4)\quad \frac{dW}{dp}=(p-c)\int_R x_p f dn - \int_R H_x x_p f dn - \int_R H_\alpha \alpha_p f dn$$

where I have omitted the arguments of $f(n)$, $x(p,n)$, $\alpha(p,n)$. The first term on the right-hand side measures the conventional deadweight loss resulting from an increase in price. Since $x_p\leqq 0$, the second term necessarily represents an offsetting decrease in health damage. The sign of the third term is in general indeterminate. But when $\alpha_p>0$, all n, the favorable health effect of a price-induced decrease in smoking frequency is counterbalanced by a compensating increase in TN. As we will see in Section IV, this is the important and realistic case.

From (4), the first-order condition for a maximum of W is

$$(5)\quad p-c=\frac{\int_R H_x x_p f dn}{\int_R x_p f dn}+\frac{\int_R H_\alpha \alpha_p f dn}{\int_R x_p f dn}$$

This condition reveals two problems in implementing a uniform corrective tax. First, we have only one policy instrument (i.e., p) to influence two different dimensions of cigarette consumption (i.e., x and α), both of which may affect health. By taxing the quantity of cigarettes consumed, we may indirectly and adversely affect individuals' TN choices (see Jerry Green and Eytan Sheshinski). Second, the variability of smokers' cigarette tastes limits the discriminatory power of a uniform tax. Individuals who smoke different quantities of cigarettes with different TN contents will in general have different marginal health damages H_x and H_α. We cannot therefore set the markup $p-c$ equal to the marginal health damage for each smoker. Given this limitation, the price derivatives x_p and α_p tell us which smokers' marginal health damages should get the most weight (see Peter Diamond).

Now define

$$(6)\quad \mu(\alpha,p,n)=\max_x \{U(x,\alpha,n)-px+Y\}$$

The indirect utility function μ represents the maximum utility achieved at a uniform price p when α is constrained at a particular value. Let $x(p,\alpha,n)$ be the maximizing choice of x in (6). We have $x_p(\alpha,p,n)\leqq 0$, and when x and α are substitutes, $x_{\alpha}(\alpha,p,n)$ is negative. Define

$$v(p,n)=\max_{\alpha}\mu(\alpha,p,n) \tag{7}$$

The indirect utility function v represents the maximum utility achieved at a uniform price p when both x and α are unconstrained. The utility-maximizing choices in (7) are necessarily $x(p,n)$ and $\alpha(p,n)$.

III. Step-Function Price Schedule

Now consider the case where different uniform tax rates apply to cigarette brands whose *TN* contents exceed or fall below a specified cutoff value. Brands with *TN* content not exceeding the cutoff value α^* have after-tax price p. Brands with *TN* content exceeding α^* have after-tax price q. The graph of such a unit price schedule, in the typical case where $q>p$, is the step function of Figure 1. The more general case of multiple cutoffs is considered in Section VI.

The representative points denoted by D, E, F in Figure 1 illustrate the three kinds of responses of individual smokers. Individuals at D smoke cigarettes with *TN* content less than α^*. Those at E, by contrast, would smoke higher *TN* cigarettes if α^* were increased. But their desires for higher *TN* are not so great that they are willing to pay the higher price q. Finally, those at F are willing to pay the higher price q for higher *TN*. Hereafter, D, E, and F refer interchangeably to subsets of points in the (α,p) plane and to subsets of smokers.

In Figure 2, indifference curves for μ for a particular smoker are superimposed upon Figure 1. The direction of increasing utility is downward. For any uniform price schedule, this smoker's choice of α is determined by the point of tangency of his indifference curve to the horizontal price line. The expansion path corresponding to these tangency points is aa'. In the presence of a step-function price schedule, however, this particular smoker's utility μ is maximized at the corner solution E. Note that $\mu^1=v(q,n)$, $\mu^2=\mu(\alpha^*,p,n)$, and $\mu^3=v(p,n)$.

The subsets D, E, and F correspond to a partitioning of the domain of n into three intervals. The boundary points between these intervals, which I denote by n' and n'', are defined as follows: n' is the root of

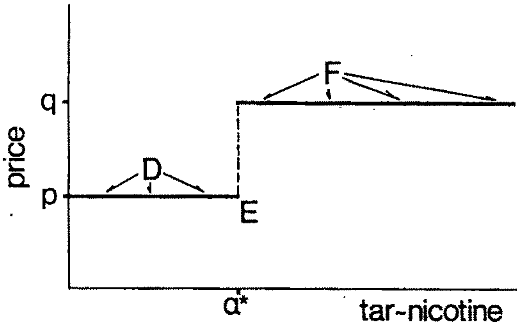


FIGURE 1. STEP-FUNCTION PRICE SCHEDULE

$$\alpha(p,n)=\alpha^* \tag{8}$$

and n'' is the root, for all $n \geqq n'$, of

$$\mu(\alpha^*,p,n)=v(q,n) \tag{9}$$

Individuals in subset D, for whom $n<n'$, smoke $x(p,n)$ cigarettes with *TN* delivery $\alpha(p,n)<\alpha^*$ and have utility $v(p,n)$. Those in subset E, for whom $n'<n<n''$, smoke $x(\alpha^*,p,n)$ cigarettes with *TN* content α^* and have utility $\mu(\alpha^*,p,n)$. Those in subset F, for whom $n>n''$, smoke $x(q,n)$ cigarettes with *TN* content $\alpha(q,n)$ and have utility $v(q,n)$. Smokers of type n' have an indifference curve tangent to the horizontal price line for p exactly at point E. Smokers of type n'' have an indifference curve that passes through point E and is tangent to the horizontal price line for q.

Can this step-function price schedule solve the two conceptual problems posed by the implementation of a uniform corrective tax? For smokers in subset E, we now have as many policy instruments (that is, p and α^*) as dimensions of cigarette consumption (i.e., x and α). In fact, we can locate an individual anywhere to the left of his aa' curve in Figure 2 by appropriately choosing α^* and $p<q$ to form a corner E at that point. (To the right of aa', we would need $q<p$, a case which will not be important below.) Because smokers' cigarette tastes are heterogeneous, however, we cannot neces-

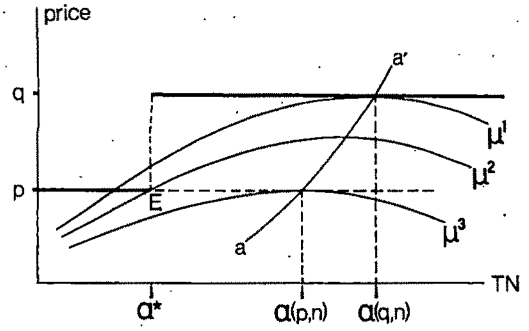


FIGURE 2. INDIFFERENCE CURVES FOR THE INDIRECT UTILITY FUNCTION μ

sarily get all smokers simultaneously to their individual optima at E. In that case, it may be preferable to bunch only some smokers at the corner E, permitting the remaining smokers in D and F to face two different uniform prices. Our regulatory design question becomes: Which cigarette smokers should end up in which subsets?

The social welfare function is

$$(10)\quad \begin{aligned} W=&\int_D\left[v(p,n)+(p-c)x(p,n)\right.\\ &\left.-H(x(p,n),\alpha(p,n))\right]f(n)dn\\ &+\int_E\left[\mu(\alpha^*,p,n)+(p-c)x(\alpha^*,p,n)\right.\\ &\left.-H(x(\alpha^*,p,n),\alpha^*)\right]f(n)dn\\ &+\int_F\left[v(q,n)+(q-c)x(q,n)\right.\\ &\left.-H(x(q,n),\alpha(q,n))\right]f(n)dn \end{aligned}$$

with partial derivatives

$$(11)\quad \begin{aligned} W_p=&(p-c)\int_{D\cup E}x_pf\,dn-\int_{D\cup E}H_xx_pf\,dn\\ &-\int_D H_\alpha\alpha_pf\,dn-(J'-J'')n_p'' \end{aligned}$$

$$(12)\quad \begin{aligned} W_q=&(q-c)\int_F x_qf\,dn-\int_F H_xx_qf\,dn\\ &-\int_F H_\alpha\alpha_qf\,dn-(J'-J'')n_q'' \end{aligned}$$

$$(13)\quad \begin{aligned} W_{\alpha^*}=&\int_E(U_\alpha+(p-c)x_\alpha)f\,dn-\int_E H_\alpha f\,dn\\ &-\int_E H_xx_\alpha f\,dn-(J'-J'')n''_{\alpha^*} \end{aligned}$$

where

$$(14)\quad \begin{aligned} J'=&\left[H(x(\alpha^*,p,n''),\alpha^*)\right.\\ &\left.-H(x(q,n''),\alpha(q,n''))\right]f(n'') \end{aligned}$$

$$(15)\quad \begin{aligned} J''=&\left[(p-c)x(\alpha^*,p,n'')\right.\\ &\left.-(q-c)x(q,n'')\right]f(n'') \end{aligned}$$

As in (4), changes in the policy parameters p, q, and α^* reflect the balancing of deadweight loss and health damage effects among the affected subsets of smokers. In contrast to (4), marginal changes in these parameters also produce discrete responses among those smokers of type n'' who jump between higher-priced higher *TN* brands in F and the lower-priced lower *TN* alternative at E. These additional welfare effects involve health damages (J'), and tax revenues and profits (J''), but not utilities. These discrete jumps cannot be dismissed as second-order magnitudes. In fact, they can dominate the other terms in the welfare gradient. Since the magnitudes of these jump effects depend on the number of smokers on the boundary between E and F (equations (14) and (15)), the welfare gradient may thus be sensitive to the underlying distribution of tastes $f(n)$. As we will see below, if $f(n)$ has many modes, then W may actually have many local optima.

The necessary first-order conditions for a welfare maximum are

$$(16)\quad p-c=\frac{\int_{D\cup E}H_xx_pf\,dn}{\int_{D\cup E}x_pf\,dn}+\frac{\int_D H_\alpha\alpha_pf\,dn}{\int_{D\cup E}x_pf\,dn}+\frac{(J'-J'')n_p''}{\int_{D\cup E}x_pf\,dn}$$

$$(17)\quad q-c=\frac{\int_F H_x x_q f\,dn}{\int_F x_q f\,dn}+\frac{\int_F H_\alpha \alpha_q f\,dn}{\int_F x_q f\,dn}+\frac{(J'-J'')n''_q}{\int_F x_q f\,dn}$$

$$(18)\quad \int_E (U_\alpha+(p-c)x_\alpha)f\,dn-J''n''_{\alpha *}=\int_E (H_x x_\alpha+H_\alpha)f\,dn-J'n''_{\alpha *}$$

where (16) and (17) are analogous to (5), and where (18) shows how the marginal benefit from an increase in α^* equals the marginal health cost.

IV. Cigarette Smoking Preferences

The analysis so far has imposed no specific form on the utility functions U. However, there is considerable scientific evidence that consumption x and *TN* delivery α are, to varying degrees, substitutes. In short-term behavioral experiments, for example, subjects increase their smoking frequency in response to controlled dilution of the cigarette smoke, and decrease their consumption frequency when cigarette tar and nicotine are increased (for example, see M. A. H. Russell et al.; Stanley Schachter; Murray Jarvik). In all of these experiments, smokers differ considerably in the magnitude of their total *TN* intake and in the extent of their compensation for changes in *TN*. Heavier smokers, it appears, compensate more for a controlled change in *TN* than light smokers. Yet no smoker appears to compensate completely for any given change in *TN* (see S. R. Sutton et al.). In general, there is no strong correlation between smoking frequency and *TN* choices across smokers (see also Section VI).

To reflect these facts, I assume that the utility functions U have the following form

$$(19)\quad U(x,\alpha,n,m)=-\frac{B}{n}(A\alpha x)^{-n}-\frac{B}{m}\alpha^{m}$$

where $1\geqq A>0$ and $B>0$ are fixed constants, and where both $n>0$ and $m\geqq 1$ are taste parameters which may vary among smokers. The unconstrained demand functions solving (3) are

$$(20)\quad x(p,n,m)=A^{mn/(m+n+mn)}\times(p/B)^{-(m+n)/(m+n+mn)}$$

$$(21)\quad \alpha(p,n,m)=A^{-n/(m+n+mn)}\times(p/B)^{n/(m+n+mn)}$$

The constrained demand function solving (6) is

$$(22)\quad x(\alpha,p,n,m)=(A\alpha)^{-n/(1+n)}(p/B)^{-1/(1+n)}$$

If $p>B$, then from (20) and (21), $x_n>0$, $\alpha_n>0$, $x_m>0$ and $\alpha_m<0$. Variations in the taste parameter n reflect heterogeneity in smokers' total *TN* preferences. Variations in the taste parameter m reflect the fact that some smokers want their total *TN* in frequent less-concentrated doses, while others want their total *TN* in infrequent more-concentrated doses.

From (21) and (22), $\alpha_p>0$ and $x_\alpha<0$, i.e., smoking frequency and *TN* content are substitutes. The elasticity of substitution between x and α (for a controlled change in α) is $\varepsilon=-(\alpha/x)x_\alpha=n/(1+n)$. Hence, the strength of substitution varies among smokers, with heavier (high n) smokers compensating for a controlled change in α more than lighter (low n) types. Since $\varepsilon<1$, no smoker will compensate completely. From (20), the unconstrained own-price elasticity of demand is $\eta=-(p/x)x_p=(m+n)/(m+n+mn)$. At any given uniform price, those who smoke a large number of cigarettes will have lower own-price elasticities.

Although Section III analyzed the case of a single taste characteristic, the step-function tax analysis of two taste parameters (n,m) is completely analogous. The subsets D, E, and F correspond to a partition of the (n,m) plane, where the boundaries between D and E and between E and F are two continuous curves. The welfare gradient is analogous to equations (11)–(15), with the corresponding jump terms J' and J'' reflecting discrete moves by smokers whose characteristics (n,m) lie on the curve separating subsets E and F.

V. Health Damage Function

The health damage due to cigarette smoking has been found to be an increasing function of both smoking frequency and *TN* content (for example, see E. Cuyler Hammond et al., 1976, 1977). The exact quantitative dose-response relations, however, have not been fully delineated.

If smoking affected only one stage of a disease process, then the disease incidence should be a linear function of total *TN* dosage. If smoking affected two independent stages, then the disease incidence should be proportional to the square of the total *TN* dosage. If the disease process itself reduced the potency of any given cigarette dosage, then the damage function would tend to be concave (for example, the loss of lung function resulting from increased smoking could diminish an individual's ability to inhale deeply.) Moreover, health damage need not depend only on total *TN* intake. In principle, smoking a large number of low *TN* cigarettes could be more or less damaging than consuming the same total *TN* in the form of a few high *TN* cigarettes (for example, the noxious effects of a particular smoke component might depend more on the peak blood level than the average exposure.) Although numerous studies have reported empirical dose-response relations between smoking frequency and disease incidence or mortality rates, their refined quantitative interpretation is known to be complicated by a variety of potential measurement biases (for example, poor health due to smoking affects smoking habits and in turn reported cigarette consumption). (See Richard Doll and Richard Peto; Hammond; Hammond et al., 1976, 1977; Peto.)

I therefore assume that the health damage function has the form

$$H(x,\alpha)=hx^{\kappa}\alpha^{\lambda} \tag{23}$$

where h, κ, and λ are positive constants. If $\kappa=\lambda$, then health damage depends only on the total *TN* intake $x\alpha$. If $\kappa>\lambda$, then smoking a large number of low *TN* cigarettes would produce more damage than obtaining the same total *TN* intake in the form of a few high *TN* cigarettes. If $\kappa<\lambda$, then a large number of low *TN* cigarettes would produce less health damage. Below I shall consider alternative values of 0.5, 1.0, and 2.0 for both κ and λ.

VI. Results

Hereafter, p is measured in dollars per pack, x is measured in packs per day, and α is measured in milligrams of tar per cigarette. Since the crude correlation coefficient between tar and nicotine contents across brands exceeds 0.9, the latter measurement convention involves little loss of generality.

Figure 3 depicts the distribution of tar per cigarette smoked by adult, current cigarette smokers in the United States in 1975. This distribution was derived from an appropriately weighted sample of 4,134 current smokers responding to the 1975 *Adult Use of Tobacco* survey. The central density in the 15–20 mg. tar range represents the conventional filter-tip brands, while the density beyond 20 mg. tar represents the conventional high *TN*, nonfilter brands. Those cigarettes below 15 mg. tar represent the relatively new low-tar and nicotine varieties. Table 1 shows the relation between cigarette tar content α and smoking frequency x among individuals in this sample. The crude correlation coefficient between x and α was 0.04.

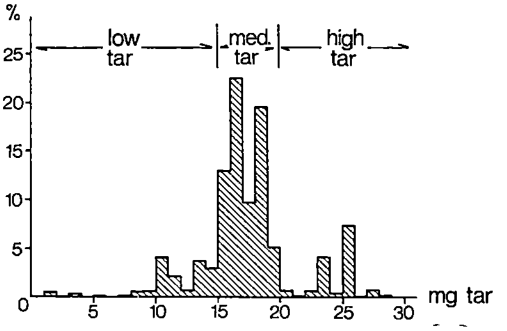


FIGURE 3. DISTRIBUTION OF TAR PER CIGARETTE; *U.S.* ADULT CURRENT SMOKERS, 1975

TABLE 1—JOINT DISTRIBUTION OF CIGARETTE TAR CONTENT AND SMOKING FREQUENCY (Shown in Percent)

Number of Packs Smoked Per Day	Cigarette Tar Content (mg.) Low Tar $\alpha \leqq 15$	Medium Tar $15<\alpha \leqq 20$	High Tar $\alpha > 20$	Total
$x \leqq 0.5$	3.8	18.3	2.7	24.8
$0.5<x \leqq 1.5$	9.3	41.8	9.0	60.1
$x>1.5$	2.4	9.8	2.9	15.1
Total	15.5	69.9	14.6	100.0

These data describe only the distribution of cigarette consumption at the uniform prices prevailing in 1975, and not the underlying distribution of taste characteristics. However, for any uniform price p, the demand functions (20) and (21) specify a one-to-one map from (n,m) to (x,α), and therefore an inverse map from the image set of (x,α) back to (n,m). We can therefore uniquely determine an individual smoker's underlying taste parameters by observing his consumption frequency and *TN* choices at a particular uniform price. When I chose values of $A=0.01$ and $B=0.001$ in (19) to (22), this procedure yielded a mean own-price elasticity η among smokers equal to 0.44 (standard deviation, 0.07; range, 0.16–0.73), and a mean elasticity of substitution ε equal to 0.76 (standard deviation, 0.10; range, 0.37–0.99). For those smoking 0.5 packs per day or less, the mean value of η was 0.54 and the mean ε was 0.63. For those smoking more than 1.5 packs per day, the corresponding means of η and ε were 0.36 and 0.87, respectively. The results derived from these choices of A and B are consistent with previous experimental studies of smoking behavior (for example, Schachter's data yield mean values of ε for light and heavy smokers equal to 0.53 and 0.74, respectively) and with econometric estimates of the own-price elasticity of demand for cigarettes (for example, see Herbert Lyon and Julian Simon; Lyon and M. Lynn Spruill; Robert Miller).

Now consider the constant h in (23). Bryan Luce and Stuart Schweitzer have estimated the total *U.S.* health damage in 1975 attributable to cigarette smoking to be $25.7 billion, including $7.5 billion in health care costs and $18.2 billion in lost earnings due to sickness and death. It is not obvious, however, what fraction of this total health damage should be counted as an unperceived cost (see Anthony Atkinson). We could assume that anyone who smokes is necessarily misinformed or shackled by addiction, and then count the entire $25.7 billion as the unperceived cost. More conservatively, we could assume that smokers ignore only those health costs subsidized by public and private insurance plans, which overwhelmingly do not distinguish smoking status or the number and type of cigarettes smoked. If about two-thirds of smoking related illness is covered by health insurance, and if about one-quarter of smoking-related lost earnings is covered by disability, pension, and life insurance plans, then the total unperceived cost in 1975 would be about $10 billion or, given 52.9 million regular smokers in 1975, a mean of $190 per smoker annually. I shall adjust h to yield this latter aggregate cost estimate.

Finally, I measure the markups $p-c$ and $q-c$ as increases in the price of cigarettes beyond observed 1975 values. In effect, I assume that current prices are those which would prevail at the optimum in the absence of unperceived health damages (see Efraim Sadka). Such an approach ignores the possibility of noncompetitive pricing by cigarette manufacturers. Because of geographic variations in cigarette excise taxes, the observed current prices in this sample ranged from $0.36 to $0.60 per pack.

Figure 4 illustrates the welfare effects of varying the cutoff α^* from 5 to 25 mg. tar, with constant taxes of $p-c=\$0.15$ and $q-c=\$0.30$, and with health damage parameters

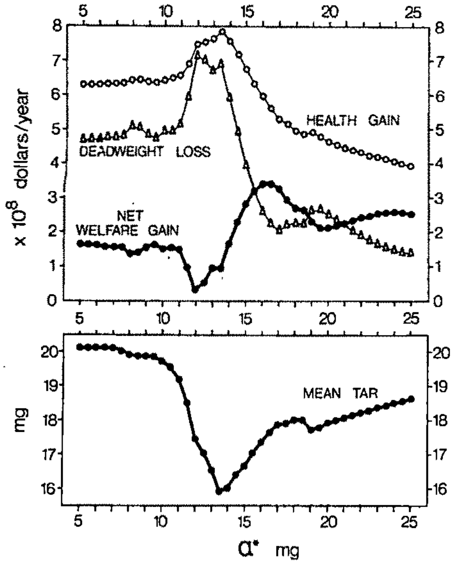


FIGURE 4. EFFECT OF VARYING α^*

$p-c=\$0.15; q-c=\$0.30; \kappa=\lambda=1$

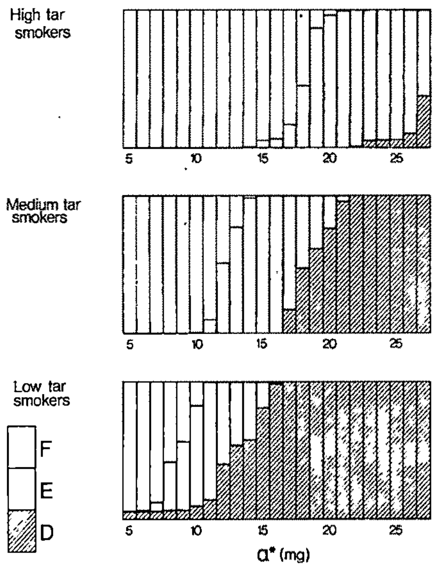


FIGURE 5. DISTRIBUTION OF LOW, MEDIUM, AND HIGH TAR SMOKERS AMONG SUBSETS D, E, AND F AS α^* VARIES

$p-c=\$0.15; q-c=\$0.30; \kappa=\lambda=1$

$\kappa=\lambda=1$. The top panel depicts the predicted net decrease in health costs, the deadweight loss, and the net welfare gain, measured in \$0.1 billions annually for the United States in 1975. The bottom panel shows the predicted mean tar per cigarette. The cutoff value $\alpha^*=13.5$ mg. yields a maximum health gain of \$0.78 billion annually, at the point of minimum average tar per cigarette. But this health gain is counterbalanced by an equally substantial deadweight loss. At the welfare maximum $\alpha^*=16.5$ mg. tar, the net welfare gain is \$0.34 billion annually. The health gain is \$0.56 billion annually, and the deadweight loss is \$0.22 billion, which represents a \$3.24 billion annual increase in tax revenues, minus a \$3.46 billion annual loss in consumer surplus.

Figure 5 displays the corresponding distribution of smokers among subsets D, E, and F. The figure depicts separately the responses of those individuals originally smoking low-tar ($\alpha \leqq 15$), medium-tar ($15<\alpha\leqq 20$), and high-tar ($\alpha>20$) cigarettes at current prices. The marked decline in mean tar as α^* increases from 10 to 14 mg. (Figure 4) corresponds to medium-tar smokers' jumping from F to E, that is, switching to lower-priced brands. At the welfare optimum, the low, medium, and high-tar smokers are almost exactly sorted between subsets D, E, and F. At this point, the medium-tar smokers reduce both x and α by about 5 percent, while the high-tar smokers reduce their smoking frequency by about 15 percent, without switching brands. As α^* increases beyond 16.5 mg. tar, the high-tar smokers jump into E. Because x and α are substitutes, their resulting decline in tar is counterbalanced by a 20 percent increase in smoking frequency.

Table 2 displays the optimal uniform and optimal step-function taxes for different values of κ and λ. As long as health damage is a function only of total TN intake (that is, $\kappa=\lambda$), the optimal cutoff α^* is 16–17 mg. tar, a value which almost exactly sorts low, medium, etc., and high-tar smokers between D, E, and F. The optimal step-function tax achieves a welfare gain twice that of the optimal uniform tax by bunching medium-tar smokers (who constitute 70 percent of smokers) at the corner E, where they decrease both their smoking frequency and cigarette tar.

TABLE 2—OPTIMAL UNIFORM AND STEP-FUNCTION TAXES

Damage Function Parameters		Optimal Uniform Tax		Optimal Step-Function Tax			
κ	λ	Tax ($/pack)	1975 Welfare Gain (billions of $)	Low-Tar Tax ($/pack)	High-Tar Tax ($/pack)	Cutoff Tar (mg.)	1975 Welfare Gain (billions of $)
0.5	0.5	0.10	0.07	0.17	0.21	16.6	0.13
	1.0	−0.07	0.04	0.18	0.78	15.0	0.49
	2.0	−0.24	1.13	0.06	0.69	12.7	2.78
1.0	0.5	0.36	0.76	0.40	0.44	18.0	0.78
	1.0	0.18	0.22	0.31	0.43	16.3	0.44
	2.0	−0.14	0.24	0.02	0.40	14.1	1.85
2.0	0.5	0.64	2.61	0.65	0.66	20.5	2.61
	1.0	0.57	1.91	0.60	0.64	19.5	1.95
	2.0	0.29	0.51	0.45	0.70	16.0	1.00

When $\kappa<\lambda$, a uniform increase in price diminishes welfare, because the health gain achieved by a price-induced decrease in consumption is completely offset by the compensating price-induced increase in *TN* (recall (4) above). The optimal step-function strategy is to impose a price differential of $q-p=\$0.38\text{–}0.63$ around a cutoff of $\alpha^*=$ 12.7 to 15.0 mg. tar. This price differential is so large that smoking brands with *TN* content above α^* is effectively prohibited. In this case, the welfare gain derives from brand switches of both medium- and high-tar smokers. (For example, at $\kappa=1$, $\lambda=2$, all medium-tar smokers and 99.6 percent of high-tar smokers are bunched at E. The total annual welfare gain of $1.85 billion reflects a $0.98 billion annual gain among medium-tar smokers and a $0.86 billion annual gain among high-tar smokers, with a negligible welfare effect on low-tar smokers.) Finally, when $\kappa>\lambda$, the step-function tax yields a net welfare gain which is indistinguishable from that achieved by a uniform tax.

Although Section III discussed only the case of two tax rates (p and q) and a single cutoff (α^*), the quantitative analysis of a step-function tax with multiple cutoffs is completely analogous. In particular, the two-cutoff case (with the population of smokers partitioned into five subsets) yielded the following results. For $\kappa=\lambda=1$, the optimal tax scheme was $0.40 per pack on brands not exceeding 16 mg. tar, $0.60 per pack on brands over 16 mg. tar but not exceeding 25 mg. tar, and $1.00 per pack on brands exceeding 25 mg. tar. The net welfare gain was $0.58 billion, as compared to $0.44 for the optimal single cutoff tax. This additional welfare gain was extracted primarily from high-tar smokers, who were induced to cut back their consumption without switching to low-tar brands. Moreover, by locating the high-tar smokers at a separate corner solution in the (α,p) plane, we can impose a higher tax on the medium-tar smokers. When $\kappa=\lambda=2$, in fact, the incremental welfare gain derived from an additional *TN* cutoff was even more substantial, because the damage function is then convex in total *TN* intake. In the $\kappa=1$, $\lambda=2$ case, however, a second *TN* cutoff yields almost no incremental welfare gain, because we already want everyone to switch to low-tar brands. When $\kappa=2$, $\lambda=1$, a second *TN* cutoff is likewise superfluous, adding little to the welfare gain achieved by a nondiscriminating uniform tax.

Would the optimal tax change if low-tar cigarettes became more popular? To test this possibility, I increased the sampling weights on all low-tar smokers in the data base and then recalculated the optimal single cutoff tax. For the $\kappa=\lambda=1$ case, increasing the effective proportion of low-tar smokers from 16 percent (Table 1) up to 62 percent had no effect on the optimal cutoff α^*. When the effective proportion of low-tar smokers exceeded 62 percent, the optimal α^* shifted to 13.4 mg. tar. Because the marginal health damages of low-tar smokers are small relative to medium- and high-tar smokers, our targeting the tax structure toward low-tar smokers yields little health gains until they reach a critical proportion of the smoking population. This critical proportion currently exceeds the projected one-third market share of low-tar cigarettes for 1979 (see my paper).

VII. Discussion and Conclusions

I have shown how the design of a *TN*-based cigarette tax hinges critically on the health tradeoff between smoking frequency and cigarette *TN* delivery. If smoking a few high-*TN* cigarettes produces more health damage than smoking many low-*TN* cigarettes, then the best policy is the prohibitively high tax on brands delivering more than 13–15 mg. tar. If smoking many low-*TN* cigarettes is more damaging, then the best policy is a relatively high, uniform tax on all brands. If health damage depends only upon total *TN* intake, then the best policy is to tax all brands, with a moderate price differential at 16–17 mg. tar. This policy induces medium-tar filter-tip smokers to reduce both *TN* and smoking frequency, while high-tar nonfilter smokers reduce their smoking frequency without switching to the lower-*TN* brands. This health tradeoff also determines the incremental welfare gains to be achieved by additional gradations in the tax scheme. These potential gains must be balanced against the administrative cost of maintaining an array of price gradations for different *TN* deliveries (see William Drayton; Donald Garner).

The annual net benefit derived from the simple step-function tax ranged from $0.1 to $2.8 billion in 1975. These results, however, were based upon a conservative interpretation of the unperceived health costs of smoking. Under a broader definition, the net welfare gains could be substantially greater. As long as the social cost of smoking is construed in terms of health damages which smokers inflict upon themselves, then the damage function can be derived from the biological dose-response curve. If the effects of sidestream smoke on nonsmokers or bandwagon effects on teenagers were at issue, however, then these dose-response relationships do not necessarily apply.

For the class of damage functions considered above, the optimal *TN* cutoff ranged from 12 to 21 mg. tar. However, if the true damage function displays a no-damage threshold at very low values of *TN*, then the optimal cutoff may fall in the very low range. At present, however, there is no evidence for the existence of such a threshold. My inclusion of only smoking frequency and cigarette *TN* delivery in the health damage function ignores other factors contributing to the health damage from smoking (for example, the synergistic effect of cigarettes and certain occupational exposures). In principle, we could write the damage function as $H(x,\alpha,r)$ where r is an additional characteristic which varies among smokers. The critical question in that case is the possible covariation between r and the taste characteristics n and m (for example, asbestos-handling smokers might be high-*TN* types). Moreover, the index α may not adequately characterize the dosage of various smoke consitituents. For example, certain medium-tar cigarettes with unventilated filters may have greater carbon monoxide deliveries than certain high-tar nonfilter cigarettes (see R. A. Jenkins, R. B. Quincy, and M. R. Guerin). In the optimal tax results above, however, high-tar smokers did not switch to these medium-tar brands and therefore did not increase their cigarette carbon monoxide content. In the $\kappa<\lambda$ case, both high-tar and medium-tar smokers switched to brands below 13–15 mg. tar, a range where carbon monoxide deliveries are generally lower (see Jenkins, Quincy, and Guerin, Figure 1). In the $\kappa>\lambda$ case, the uniform tax induced some medium-tar smokers to switch to high-tar nonfilter brands, thus decreasing their carbon monoxide content.

On the supply side, this analysis assumed that the marginal cost of cigarettes was constant and independent of *TN*. Accurate information concerning the allocation of various tobacco manufacturing, product development, and promotional costs across brands is not currently available. The effect of alternative tax schemes on the magnitude and distribution of cigarette advertising expenditures is a critical issue in this respect.

On the demand side, this analysis omitted certain aspects of cigarette consumption other than brand choice and smoking frequency. When smokers switch to lower *TN* cigarettes, for example, they may compensate by inhaling more deeply or smoking them further down to the end (see, for example, Sutton et al.; the author). It is difficult to imagine how we could devise an additional policy instrument to control this aspect of consumption. Moreover, by assuming a fixed population of smokers, I have ignored the possibility that changes in taxes may affect initiation of smoking among teenagers or cessation of smoking among adults.

Finally, this analysis has not squarely confronted the addictive nature of smoking. Although my welfare criterion incorporates the consumer benefit from smoking, it may not be appropriate to measure this benefit in terms of willingness to pay. This is the subject of a later paper.

REFERENCES

A. B. Atkinson, "Smoking and the Economics of Government Intervention," in Mark Perlman, ed., *The Economics of Health and Medical Care*, New York 1973.

P. A. Diamond, "Consumption Externalities and Imperfect Corrective Pricing," *Bell. J. Econ.*, Autumn 1973, *4*, 526–38.

R. Doll and R. Peto, "Cigarette Smoking and Bronchial Carcinoma: Dose and Time Relationships Among Regular Smokers and Lifelong Non-Smokers," *J. Epidemiol. Commun. Health*, Dec. 1978, *32*, 303–13.

W. Drayton Jr., "The Tar and Nicotine Tax: Pursuing Public Health Through Tax Incentives," *Yale Law. J.*, July 1972, *81*, 1487–516.

D. W. Garner, "Cigarettes and Welfare Reform," *Emory Law J.*, 1977, *26*, 269–335.

J. Green and E. Sheshinski, "Direct versus Indirect Remedies for Externalities," *J. Polit. Econ.*, July/Aug. 1976, *84*, 797–808.

E. C. Hammond, "Quantitative Relationship Between Cigarette Smoking and Death Rates," *Nat. Cancer Inst. Monog.*, 1968, *28*, 3–8.

______ **et al.**, " 'Tar' and Nicotine Content of Cigarette Smoke in Relation to Death Rates," *Environ. Res.*, Dec. 1976, *12*, 263–74.

______ **et al.**, "Some Recent Findings Concerning Cigarette Smoking," in H. H. Hiatt et al., eds., *Origins of Human Cancer*, Cold Spring Harbor 1977.

J. E. Harris, "Cigarette Smoking in the United States, 1950–1978," in *Smoking and Health: A Report of the Surgeon General*, Washington 1979.

M. E. Jarvik, "Biological Factors Underlying the Smoking Habit," in Murray E. Jarvik et al., eds., *Research on Smoking Behavior*, Nat. Inst. on Drug Abuse, Monog. 17, Washington 1977.

R. A. Jenkins, R. B. Quincy, and M. R. Guerin, "Selected Constituents in the Smokes of U.S. Commercial Cigarettes: 'Tar,' Nicotine and Carbon Monoxide," tech. memo. no. 6870, Oak Ridge National Laboratory, Apr. 1979.

B. R. Luce and S. O. Schweitzer, "The Economic Costs of Smoking-Induced Illness," in Murray E. Jarvik et al., eds., *Research on Smoking Behavior*, Nat. Inst. on Drug Abuse, Monog. 17, Washington 1977.

H. L. Lyon and J. L. Simon, "Price Elasticity of Demand for Cigarettes in the United States," *Amer. J. Agr. Econ.*, Nov. 1968, *50*, 888–95.

______ **and M. L. Spruill**, "A Temporal Cross-Section Analysis of Cigarette Price Elasticity in the United States," work. paper no. WP(1977)-5, Univ. Houston, Sept. 1977.

R. H. Miller, "Pricing Out Tobacco: Price as a Factor in Cigarette Consumption," U.S. Department of Agriculture, Washington, June 1975.

R. Peto, "Epidemiology, Multistage Models, and Short-Term Mutagenicity Tests," in

H. H. Hiatt et al., eds., *Origins of Human Cancer*, Cold Spring Harbor 1977.

M. A. H. Russell et al., "Comparison of Effect on Tobacco Consumption and Carbon Monoxide Absorption of Changing to High and Low Nicotine Cigarettes," *British Med. J.*, Dec. 1973, *4*, 512–16.

E. Sadka, "On the Optimal Taxation of Consumption Externalities," *Quart. J. Econ.*, Feb. 1978, *366*, 165–74.

S. Schachter, "Nicotine Regulation in Light and Heavy Smokers," *J. Exp. Psych. (Gen.)*, Mar. 1977, *106*, 5–12.

S. R. Sutton et al., "Adjustment of Smokers to Dilution of Tobacco Smoke by Ventilated Cigarette Holders," *Clin. Pharmacol. Ther.*, Oct. 1978, *24*, 395–405.

U.S. Department of Health, Education, and Welfare, *Adult Use of Tobacco 1975*, Atlanta 1976.

An Almost Ideal Demand System

By ANGUS DEATON AND JOHN MUELLBAUER*

Ever since Richard Stone (1954) first estimated a system of demand equations derived explicitly from consumer theory, there has been a continuing search for alternative specifications and functional forms. Many models have been proposed, but perhaps the most important in current use, apart from the original linear expenditure system, are the Rotterdam model (see Henri Theil, 1965, 1976; Anton Barten) and the translog model (see Laurits Christensen, Dale Jorgenson, and Lawrence Lau; Jorgenson and Lau). Both of these models have been extensively estimated and have, in addition, been used to test the homogeneity and symmetry restrictions of demand theory. In this paper, we propose and estimate a new model which is of comparable generality to the Rotterdam and translog models but which has considerable advantages over both. Our model, which we call the Almost Ideal Demand System (*AIDS*), gives an arbitrary first-order approximation to any demand system; it satisfies the axioms of choice exactly; it aggregates perfectly over consumers without invoking parallel linear Engel curves; it has a functional form which is consistent with known household-budget data; it is simple to estimate, largely avoiding the need for non-linear estimation; and it can be used to test the restrictions of homogeneity and symmetry through linear restrictions on fixed parameters. Although many of these desirable properties are possessed by one or other of the Rotterdam or translog models, neither possesses all of them simultaneously.

In Section I of the paper, we discuss the theoretical specification of the *AIDS* and justify the claims in the previous paragraph. In Section II, the model is estimated on postwar British data and we use our results to test the homogeneity and symmetry restrictions. Our results are consistent with earlier findings in that both sets of restrictions are decisively rejected. We also find that imposition of homogeneity generates positive serial correlation in the errors of those equations which reject the restrictions most strongly; this suggests that the now standard rejection of homogeneity in demand analysis may be due to insufficient attention to the dynamic aspects of consumer behavior. Finally, in Section III, we offer a summary and conclusions. We believe that the results of this paper suggest that the *AIDS* is to be recommended as a vehicle for testing, extending, and improving conventional demand analysis. This does not imply that the system, particularly in its simple static form, is to be regarded as a fully satisfactory explanation of consumers' behavior. Indeed, by proposing a demand system which is superior to its predecessors, we hope to be able to reveal more clearly the problems and potential solutions associated with the usual approach.

I. Specification of the *AIDS*

In much of the recent literature on systems of demand equations, the starting point has been the specification of a function which is general enough to act as a second-order approximation to any arbitrary direct or indirect utility function or, more rarely, a cost function. For examples, see Christensen, Jorgenson, and Lau; W. Erwin Diewert (1971); or Ernst Berndt, Masako Darrough, and Diewert. Alternatively, it is possible to use a first-order approximation to the demand functions themselves as in the Rotterdam model, see Theil (1965, 1976); Barten. We shall follow these approaches in terms of generality but we start, not from some arbitrary preference ordering, but from a specific class of preferences, which by the theorems of Muellbauer (1975, 1976) permit exact aggregation over consumers: the representation of market demands as if they were the outcome of decisions by a rational representative consumer. These preferences, known as the *PIGLOG* class, are represented via the *cost* or *expenditure function* which defines the minimum expenditure necessary to attain a specific utility level at given prices. We denote this function $c(u,p)$ for utility u and price vector p, and define the *PIGLOG* class by

---

*University of Bristol, and Birkbeck College, London, respectively. We are grateful to David Mitchell for help with the calculations and to Anton Barten, David Hendry, Claus Leser, Louis Phlips, and a referee for helpful comments on an earlier version.

$$\log c(u,p) = (1-u)\log\{a(p)\} + u\log\{b(p)\} \tag{1}$$

With some exceptions (see the Appendix), u lies between 0 (subsistence) and 1 (bliss) so that the positive linearly homogeneous functions $a(p)$ and $b(p)$ can be regarded as the costs of subsistence and bliss, respectively. The Appendix further discusses this general model as well as the implications of the underlying aggregation theory.

Next we take specific functional forms for $\log a(p)$ and $\log b(p)$. For the resulting cost function to be a flexible functional form, it must possess enough parameters so that at any single point its derivatives $\partial c/\partial p_i$, $\partial c/\partial u$, $\partial^2 c/\partial p_i \partial p_j$, $\partial^2 c/\partial u \partial p_i$, and $\partial^2 c/\partial u^2$ can be set equal to those of an arbitrary cost function. We take

$$\log a(p) = a_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_j \gamma^*_{kj} \log p_k \log p_j \tag{2}$$

$$\log b(p) = \log a(p) + \beta_0 \prod_k p_k^{\beta_k} \tag{3}$$

so that the *AIDS* cost function is written

$$\log c(u,p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_j \gamma^*_{kj} \log p_k \log p_j + u\beta_0 \prod_k p_k^{\beta_k} \tag{4}$$

where α_i, β_i, and γ^*_{ij} are parameters. It can easily be checked that $c(u,p)$ is linearly homogeneous in p (as it must be to be a valid representation of preferences) provided that $\Sigma_i \alpha_i = 1, \Sigma_j \gamma^*_{kj} = \Sigma_k \gamma^*_{kj} = \Sigma_j \beta_j = 0$. It is also straightforward to check that (4) has enough parameters for it to be a flexible functional form provided it is borne in mind that, since utility is ordinal, we can always choose a normalization such that, at a point, $\partial^2 \log c/\partial u^2 = 0$. The choice of the functions $a(p)$ and $b(p)$ in (2) and (3) is governed partly by the need for a flexible functional form. However, the main justification is that this particular choice leads to a system of demand functions with the desirable properties which we demonstrate below.

The demand functions can be derived directly from equation (4). It is a fundamental property of the cost function (see Ronald Shephard, 1953, 1970, or Diewert's 1974 survey paper) that its price derivatives are the quantities demanded: $\partial c(u,p)/\partial p_i = q_i$. Multiplying both sides by $p_i/c(u,p)$ we find

$$\frac{\partial \log c(u,p)}{\partial \log p_i} = \frac{p_i q_i}{c(u,p)} = w_i \tag{5}$$

where w_i is the budget share of good i. Hence, logarithmic differentiation of (4) gives the budget shares as a function of prices and utility:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i u \beta_0 \prod p_k^{\beta_k} \tag{6}$$

where

$$\gamma_{ij} = \frac{1}{2}(\gamma^*_{ij} + \gamma^*_{ji}) \tag{7}$$

For a utility-maximizing consumer, total expenditure x is equal to $c(u,p)$ and this equality can be inverted to give u as a function of p and x, the indirect utility function. If we do this for (4) and substitute the result into (6) we have the budget shares as a function of p and x; these are the *AIDS* demand functions in budget share form:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\{x/P\} \tag{8}$$

where P is a price index defined by

$$(9)\qquad \log P = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_j \sum_k \gamma_{kj} \log p_k \log p_j$$

The restrictions on the parameters of (4) plus equation (7) imply restrictions on the parameters of the *AIDS* equation (8). We take these in three sets

$$(10)\qquad \sum_{i=1}^{n} \alpha_i = 1 \qquad \sum_{i=1}^{n} \gamma_{ij} = 0 \qquad \sum_{i=1}^{n} \beta_i = 0$$

$$(11)\qquad \sum_j \gamma_{ij} = 0$$

$$(12)\qquad \gamma_{ij} = \gamma_{ji}$$

Provided (10), (11), and (12) hold, equation (8) represents a system of demand functions which add up to total expenditure ($\Sigma w_i = 1$), are homogeneous of degree zero in prices and total expenditure taken together, and which satisfy Slutsky symmetry. Given these, the *AIDS* is simply interpreted: in the absence of changes in relative prices and "real" expenditure (x/P) the budget shares are constant and this is the natural starting point for predictions using the model. Changes in relative prices work through the terms γ_{ij}; each γ_{ij} represents 10^2 times the effect on the ith budget share of a 1 percent increase in the jth price with (x/P) held constant. Changes in real expenditure operate through the β_i coefficients; these add to zero and are positive for luxuries and negative for necessities. Further interpretation is best done in terms of the claims made in the introduction.

A. *Aggregation Over Households*

The aggregation theory developed in Muellbauer (1975, 1976, of which the main relevant points are summarized in the Appendix) implies that exact aggregation is possible if, for an individual household h, behavior is described by the generalization of (8):

$$(8')\qquad w_{ih} = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\{x_h / k_h P\}$$

The parameters k_h can be interpreted as a sophisticated measure of household size which, in principle, could take account of age composition, other household characteristics, and economies of household size; and which is used to deflate the budget x_h to bring it to a "needs corrected" *per capita* level. This allows a limited amount of taste variation across households. The share of aggregate expenditure on good i in the aggregate budget of all households, denoted $\bar{w}_i$ is given by

$$\sum_h p_i q_{ih} \Big/ \sum x_h \equiv \sum_h x_h w_{ih} \Big/ \sum x_h$$

which, when applied to (8′) gives

$$(8'')\qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log p_j - \beta_i \log P + \beta_i \left\{ \sum_h x_h \log(x_h / k_h) \Big/ \sum x_h \right\}$$

Define the aggregate index k by

$$(13)\qquad \log(\bar{x}/k) \equiv \sum_h x_h \log(x_h / k_h) \Big/ \sum x_h$$

where $\bar{x}$ is the average level of total expenditure x_h. Hence (8″) becomes

$$(8''')\qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(\bar{x}/kP)$$

This is identical in form to (8′) and this confirms that under these assumptions aggregate budget shares correspond to the decisions of a rational representative household whose preferences are given by the *AIDS* cost function (4) and whose budget is given by $\bar{x}/k$, the "representative budget level."

The index k has an interesting interpretation. If each household had the same tastes ($k_h = 1$, all h), k would be an index of the equality of the distribution of household budgets. In fact, this index is identical to Theil's (1972) entropy measure of equality Z deflated by the number of households H, where $\log Z = -\Sigma(x_h/X)\log(x_h/X)$ and X is the aggregate budget; Z reaches its maximum level of H when there is perfect equality so that $x_h = \bar{x}$, all h. Therefore as inequality increases, $k = Z/H$ decreases and the representative budget level increases. When k_h differs across households, for example, because of differences in household composition, the index k reflects not only the distribution of budgets but the demographic structure. Ideally, one might attempt to model the variation of k_h with household characteristics in a cross-section study and, given time-series data on the joint distribution of household budgets and characteristics, construct a series for k for use in fitting (8‴). Data limitations have prevented us from carrying out this proposal in the empirical application below. To the extent that k is constant or uncorrelated with $\bar{x}$ or p, no omitted variable bias arises from our procedure of omitting k and redefining $\alpha_i^* = \alpha_i - \beta_i \log k^*$ where k^* is the constant or sample mean value of k.

When the distribution of household budgets and household characteristics is invariant except for equiproportional changes in household budgets, k is constant. In this case there is considerable extra scope for taste variations in the individual demand functions without altering the validity of the representative consumer hypothesis embodied in (8‴). Indeed, it turns out that not only α_{ih}, all i, but also γ_{ijh}, all i,j, can differ over households. The α_i and γ_{ij} parameters in (8‴) are then weighted averages of the micro parameters.

B. *Generality of the Model*

The flexible functional form property of the *AIDS* cost function implies that the demand functions derived from it are first-order approximations to any set of demand functions derived from utility-maximizing behavior. The *AIDS* is thus as general as other flexible forms such as the translog or the Rotterdam models. However, if maximizing behavior is not assumed but it is simply held that demands are continuous functions of the budget and of prices, then the *AIDS* demand functions (8) (without the restrictions (11) and (12)) can still provide a first-order approximation. In general, without maximizing assumptions, we can think of the budget shares w_i as being unknown functions of $\log p$ and $\log x$. From (8) and (9), the *AIDS* has derivatives $\partial w_i/\partial \log x = \beta_i$ and $\partial w_i/\partial \log p_j = \gamma_{ij} - \beta_i\alpha_j - \beta_i\Sigma\gamma_{jk}\log p_k$ so that, at any point, β and γ can be chosen so that the derivatives of the *AIDS* will be identical to those of any true model. Given that the α parameters act as intercepts, the *AIDS* can thus provide a local first-order approximation to any true demand system, whether derived from the theory of choice or not. This property is important since it means that tests of homogeneity of symmetry are set within a maintained hypothesis which makes sense and would be widely accepted in its own right.

Generality is not without its problems, however. There is a large number of parameters in (18) and on most data sets these are unlikely to be all well determined. It is thus important that there should exist some straightforward procedure for eliminating unnecessary parameters without untoward consequences for the properties of the model. In the *AIDS*, this can be done by placing whatever restrictions on γ_{ij} parameters are thought to be empirically or theoretically plausible. As we shall see below, in many cases it will be possible to impose these restrictions on a single equation basis. One obvious restriction is that for some pairs $(i,j), \gamma_{ij}$ should be zero; for such pairs the budget share of each is independent of the price of the other if (x/P) is held constant. It can be shown that γ_{ij} has approximately the same sign as the compensated cross-price elasticity and this is also useful in suggesting prior restrictions. We should not however expect *all* the γ_{ij}s to be zero; the resulting model, $w_i = \alpha_i + \beta_i \log(x/P)$ is extremely restrictive and has been tested and rejected by Deaton (1978).

C. *Restrictions*

If we start from equations (8) and (9) as our maintained hypothesis, we can examine the effects of the restrictions (10)–(12) which are required to make the model consistent with the theory of demand. The conditions (10) are the *adding-up* restrictions; as can easily be checked from (8), these ensure that $\Sigma w_i \equiv 1$. *Homogeneity* of the demand functions requires restriction (11) which can be tested equation by equation. Slutsky *symmetry* is satisfied by (8) if and only if the symmetry restriction (12) holds. As is true of other flexible functional forms, *negativity* cannot be ensured by any restrictions on the parameters alone. It can however be checked for any given estimates by calculating the eigenvalues of the Slutsky matrix s_{ij}, say. In practice, it is easier to use not s_{ij} but $k_{ij} = p_i p_j s_{ij}/x$, the eigenvalues of which have the same signs as those of s_{ij} and which are given by

$$(14) \quad k_{ij} = \gamma_{ij} + \beta_i \beta_j \log \frac{x}{P} - w_i \delta_{ij} + w_i w_j$$

where δ_{ij} is the Kronecker *delta*. Note that apart from this negativity condition, all the restrictions are expressible as linear constraints involving only the parameters and so can be imposed globally by standard techniques.

D. *Estimation*

In general, estimation can be carried out by substituting (9) in (8) to give

$$(15) \quad w_i = (\alpha_i - \beta_i \alpha_0) + \sum_j \gamma_{ij} \log p_j + \beta_i \left\{ \log x - \sum_k \alpha_k \log p_k - \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j \right\}$$

and estimating this non-linear system of equations by maximum likelihood or other methods with and without the restrictions (11) and (12). (Note that since the data add up by construction, (10) is not testable.) Equation (15) is not particularly difficult to estimate since the first-order conditions for likelihood maximization are linear in α and γ given β and vice versa so that "concentration" allows iteration on a subset of the parameters (see for example, Deaton, 1975, pp. 46–49). Although all the parameters in (15) are identified given sufficient variation in the independent variables, in many examples the practical identification of α_0 is likely to be problematical. This parameter is only identified from the α_is in (15) by the presence of these latter inside the term in braces, originally in the formula for *log P*, equation (9). However, in situations where individual prices are closely collinear, *log P* is unlikely to be very sensitive to its weights so that changes in the intercept term in (15) due to variations in α_0 can be offset in the αs with minimal effect on *log P*. This can be overcome in practice by assigning a value to α_0 a priori. Since the parameter can be interpreted as the outlay required for a minimal standard of living when prices are unity (usually in the base year; see the Appendix), choosing a plausible value is not difficult.

However, in many situations, it is possible to exploit the collinearity of the prices to yield a much simpler estimation technique. Note from (8) that if P were known, the model would be linear in the parameters α, β, and γ, and estimation (at least without cross-equation restrictions such as symmetry) can be done equation by equation by *OLS* which, in this case and given normally distributed errors, is equivalent to maximum likelihood estimation for the system as a whole. The adding-up constraints (10) will be automatically satisfied by these estimates. In situations where prices are closely collinear, it may well be adequate to *approximate* P as proportional to some known index P^*, say. One obvious candidate in view of (8) and (9) is Stone's (1953) index $\log P^* = \Sigma w_k \log p_k$. If $P \simeq \phi P^*$ say, then (8) can be estimated as

$$(16) \quad w_i = (\alpha_i - \beta_i \log \phi) + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{x}{P^*}\right)$$

Note that in this framework the α_i parameters are identified only up to a scalar multiple of β_i; if we write $\alpha_i^* = \alpha_i - \beta_i \log \phi$, it is easily seen that $\Sigma\alpha_k^* = 0$ is still required for adding up, since $\Sigma\beta_k = 0$.

In the empirical results below we shall estimate both (15) and (16), and show that the latter is an excellent approximation to the former. However, it must be emphasized that (16) exists only as an approximation to (15) and will only be accurate in specific circumstances, albeit widely occurring ones in time-series estimation. Note finally that if single equation estimation is used to investigate likely restrictions amongst the γ parameters, as is suggested in Part B above, the *constrained OLS* estimates will no longer automatically be maximum likelihood, efficient, or satisfy adding up. Hence, once the restrictions have been selected, (15) should be used to reestimate the whole system.

E. *Relationship with Budget Studies and with the Rotterdam Model*

The Engel curves corresponding to (8) take the form $p_i q_i = \xi_i x + \beta_i x \log x$ for appropriate functions of prices ξ_i. These are clearly not linear except in the proportional case when $\beta_i = 0$. The model thus allows a possible reconciliation between time-series models, which have to date required linearity of Engel curves for aggregation with cross-section results, which typically find evidence of nonlinearity. Indeed, the *PIGLOG* Engel curve $w_i = \xi_i + \beta_i \log x$ was used as early as 1943 by Holbrook Working and has recently been recommended by Claus Leser (1963, 1976) as providing an excellent fit to cross-section data in a wide range of circumstances.

In the time-series context, the *AIDS* has a close relationship to the Rotterdam model of Theil (1965, 1976) and Barten. The first difference form of (8) is

$$(17)\quad \Delta w_i = \beta_i \Delta \log\left(\frac{x}{P}\right) + \sum_j \gamma_{ij} \Delta \log p_j$$

which no longer involves the α parameters except through $\Delta \log P$. This dependence can be seen by writing (17) in full, i.e.,

$$(18)\quad \Delta w_i = \beta_i \left\{ \Delta \log x - \sum_k \alpha_k \Delta \log p_k - \frac{1}{2} \sum_k \sum_j \gamma_{kj} \Delta(\log p_j \log p_k) \right\} + \sum_j \gamma_{ij} \Delta \log p_j$$

Again, all the parameters (except α_0) are theoretically identified, but in practice the substitutability between γ_{ij}s and $\beta_i \alpha_j$s in fitting (18) if prices are nearly collinear means that, in such cases, the only practical way of estimating (17) is to replace $\Delta \log P$ by some index, for example, $\Delta(\Sigma w_k \log p_k)$ as before or by its approximation $\Sigma w_k \Delta \log p_k$. In the latter case, the right-hand side of (17) becomes identical to the right-hand side of the Rotterdam model which is

$$(19)\quad w_i \Delta \log q_i = b_i \{\Delta \log x - \Sigma w_k \Delta \log p_k\} + \sum_j c_{ij} \Delta \log p_j$$

The dependent variable is different in the *AIDS*; instead of $w_i \Delta \log q_i$ we have Δw_i or $w_i \Delta \log w_i$. Thus, by replacing the dependent variable $w_i \Delta \log q_i$ in the Rotterdam model by $w_i \Delta \log w_i$, an addition of $w_i \Delta \log(p_i/x)$, we generate the first-difference form of the *AIDS*. The similarity between the two models is quite striking in this form; both are effectively linear and both can be used to test homogeneity and symmetry with only linear restrictions on constant parameters. Note however that the parameters have quite different interpretations in the two models so that, for example, the negativity condition applies directly to the matrix of price effects in the Rotterdam model which is not the case for the *AIDS*. The crucial difference between the two models is that (17), unlike (19), is derived from explicit demand functions, (8), and an explicit characterization of preferences, (4). For the prediction of demand this difference may not be vitally important, but in many other contexts, for example in calculating cost-of-living indices, household equivalence scales, or optimal tax rates, the ability to link estimated parameter values to preferences themselves becomes of great significance.

II. An Application to Postwar British Data

In this section we estimate the model using annual British data from 1954 to 1974 inclusive on eight nondurable groups of consumers' expenditure, namely, food, clothing, housing services, fuel, drink and tobacco, transport and communication services, other goods, and other services. As discussed in Section I, Part A above, if we assume that the index k in (8‴) is either constant or that its deviations are independently distributed from those of the average budget $\bar{x}$ and of prices, no biases result from its omission. In particular, we allow the intercepts in (8‴) to absorb the $-\beta_i \log k$ terms. We then proceed by first following the strategy outlined in Section I, Part D, setting $\log P^* = \Sigma w_k \log p_k$ for each year and estimating equation (16) for each good separately by *OLS*. The system is then reestimated, equation by equation, and again using P^*, in order to test the homogeneity condition. Equation (11) is imposed by substitution so that instead of (16), we estimate

$$(20) \quad \bar{w}_i = \alpha_i^* + \sum_{j=1}^{n-1} \gamma_{ij} \log\left(\frac{p_j}{p_n}\right) + \beta_i \log\left(\frac{\bar{x}}{P^*}\right)$$

At this stage, F-ratios are calculated equation by equation to test the validity of the restriction.

The next stage is to impose symmetry of γ, at this point replacing P^* by the "correct" price index (9) with α_0 set to some appropriate value. Since symmetry, unlike homogeneity or the unrestricted model, involves cross-equation restrictions, the variance-covariance matrix of the residuals for the first time plays a part in the estimation. Since this is unknown a priori, normal practice would be to replace it by its maximum likelihood estimate. However, with only twenty-one observations, this is not practicable for equation (15) since, with so many parameters in each equation, the likelihood can be made arbitrarily large by making any *one* equation fit perfectly.[1] This difficulty can only be resolved by assuming a particular structure for the variance-covariance matrix of the residuals. Following Deaton (1975, p. 39), we assume $V = \sigma^2(I - ii')$, where V is the variance-covariance matrix of the residuals, σ^2 is a (positive) parameter to be estimated, I is an $n \times n$ identity matrix and i is a vector each of the elements of which is $(n)^{-1/2}$. In this case, maximum-likelihood estimation reduces to least squares so that instead of minimizing the *determinant* of the matrix of residual cross products, we minimize its *trace*. The likelihood values quoted below are calculated on this assumption. Once again, maximum use is made of substitution in the estimation so that, under symmetry, (15) is estimated using only the fourteen independent α_i^*s and β_is and the twenty-eight parameters forming the upper right-hand triangle of γ with its final row and column deleted. We now check that P^* and P are sufficiently close to allow comparison of likelihoods both by direct evaluation of both indices and by reestimation of the unrestricted and homogeneous models using P as evaluated from the symmetric estimates. It is also possible to check concavity at this stage by using the symmetric parameter estimates to calculate the eigenvalues of the matrix in (14). Finally, the whole process is repeated with the model written in first differences, that is, equation (17) with the addition of intercepts. Collinearity prevented any successful attempt to link P to the parameter estimates in these regressions; instead, the value of P calculated from the symmetric estimation in levels was used throughout.

Note that we choose to test symmetry whether or not homogeneity is rejected. This procedure has been criticized by Grayham Mizon who suggests that optimal inference requires that further testing be abandoned as soon as a rejection is encountered. Mizon's criticism would be correct if we were certain of our maintained hypothesis, but to some extent this is a matter of choice. Many economists would choose not to test

---

[1] We are grateful to Teun Kloek for pointing this out to us.

TABLE 1—THE UNCONSTRAINED PARAMETER ESTIMATES AND TESTS OF HOMOGENEITY
(t-Values in Parentheses)

Commodity i	α_i^*	β_i	γ_{i1}	γ_{i2}	γ_{i3}	γ_{i4}	γ_{i5}	γ_{i6}	γ_{i7}	γ_{i8}	$\sum_j \gamma_{ij}$	*S.E.E.* (10^{-2})	R^2	*D.W.*
Food	1.221	−0160	0.186	−0.077	−0.013	−0.020	−0.058	0.032	0.015	−0.098	−0.033	.113	0.999	2.33
	(7.4)	(−6.1)	(9.8)	(−4.3)	(−0.8)	(−1.1)	(−6.2)	(1.3)	(0.7)	(−4.2)	(−4.4)	.180	0.998	1.74
Clothing	−0.482	0.091	0.033	0.016	−0.024	−0.026	−0.029	0.014	0.033	−0.049	−0.032	.106	0.984	2.29
	(−3.1)	(3.7)	(1.8)	(1.0)	(−1.6)	(−1.5)	(−3.3)	(0.6)	(1.6)	(−2.2)	(−4.5)	.171	0.955	1.55
Housing	0.793	−0.104	−0.082	−0.9	0.088	0.9	0.033	−0.055	−0.030	0.098	0.051	.086	0.999	1.89
	(6.3)	(−5.1)	(−5.6)	(−0.7)	(7.2)	(0.7)	(4.7)	(−2.9)	(−1.8)	(5.5)	(9.1)	.241	0.992	1.29
Fuel	−0.159	0.033	−0.042	0.010	−0.011	0.037	−0.004	0.022	0.007	−0.031	−0.010	.140	0.883	2.25
	(−0.8)	(1.0)	(−1.8)	(0.4)	(−0.5)	(1.6)	(−0.3)	(0.7)	(0.3)	(−1.1)	(−1.1)	.141	0.870	2.03
Drink and	−0.043	0.028	−0.043	0.034	−0.027	−0.020	0.056	0.005	−0.018	0.014	0.001	.099	0.969	2.96
Tobacco	(−0.3)	(1.2)	(−2.6)	(2.2)	(−1.9)	(−1.2)	(6.9)	(0.2)	(−0.9)	(0.7)	(0.0)	.095	0.969	2.93
Transport and	−0.061	0.029	−0.022	−0.012	−0.002	0.011	0.060	−0.023	−0.024	0.053	0.040	.047	1.000	2.24
Communication	(−0.9)	(2.6)	(−2.7)	(−1.6)	(−0.3)	(1.4)	(15.2)	(−2.2)	(−2.6)	(5.3)	(13.1)	.184	0.992	1.36
Other Goods	−0.038	0.022	0.001	−0.003	−0.001	−0.006	−0.030	0.007	0.032	−0.006	−0.005	.108	0.885	1.92
	(−0.2)	(0.9)	(0.0)	(−0.2)	(−0.0)	(−0.3)	(−3.4)	(0.3)	(1.5)	(−0.2)	(0.7)	.106	0.880	1.91
Other Services	−0.231	0.060	−0.032	0.041	−0.011	0.014	−0.028	−0.003	−0.015	0.019	−0.014	.107	0.843	2.27
	(−1.5)	(2.4)	(−1.8)	(2.4)	(−0.7)	(0.8)	(−3.1)	(−0.1)	(−0.7)	(0.9)	(−2.0)	.119	0.788	1.98

homogeneity, treating absence of money illusion as a maintained hypothesis; the test of symmetry would then be the interesting one. Even if the maintained hypothesis turns out to be false, tests based on it are not necessarily without interest. Few if any tests in econometrics are carried out within the framework of maintained hypotheses which are even widely accepted, let alone of unchallengeable validity.

Table 1 reports the first-stage estimates of (16) using P^* and without any constraints on the parameters save (10) which are automatically and costlessly satisfied. The estimates of β classify food and housing as necessities while the other goods are luxuries. A large number of γ coefficients are significantly different from zero; twenty-two out of sixty-four have t-values absolutely larger than 2. Even so, none of the variables considered have any detectable effect on the value share for fuel and very few have influence in the other goods or other services equations. Similarly, the prices of fuel, of transport and communication, and of other services have little or no effect anywhere (except, of course, through P^* and the value share itself), while the prices of food, drink and tobacco, and of other services appear with considerable regularity. The total expenditure and own-price elasticities are shown in the first two columns of Table 2 and, although food has an (insignificant) positive price elasticity, these numbers appear both credible and in line with other studies. Note the general price inelasticity of demand; only transport and communication appear to be price elastic.

Table 1 also shows, in the column headed $\Sigma\gamma_{ij}$, the row sums of the unconstrained γ_{ij} matrix; this number shows 10^2 times the absolute effect on each value share of a 1 percent increase in all prices and total expenditure. Under homogeneity, this should be zero and the bracketed numbers given are t-tests of the significance of the deviation from zero. These numbers are, of course, identical to the square roots of the F-ratios obtained by comparing the residual sums of squares of equations (16) and (20). Hence, a proportional increase in prices and expenditure will decrease expenditure on food and on clothing, and increase expenditure on housing and transport and communication. These are also the commodities for which the elasticities suffer the largest changes between columns 1 and 2 and columns 3 and 4 of Table 2. Other deviations from homogeneity appear not to be significant. The final columns of Table 1 give equations standard errors, the R^2 and Durbin-Watson ($D.W.$) statistics for free and restricted estimation. Note that for the four commodity groups where homogeneity

TABLE 2—TOTAL EXPENDITURE AND OWN-PRICE ELASTICITES

	Levels Model				First-Differences Model			
	Unconstrained		Homogeneous		Unconstrained		Homogeneous	
	e_i	e_{ii}	e_i	e_{ii}	e_i	e_{ii}	e_i	e_{ii}
Food	0.21	0.07	0.04	−0.01	0.04	0.22	0.17	−0.00
Clothing	2.00	−0.92	1.51	−0.48	2.83	−0.94	2.92	−0.94
Housing	0.30	−0.31	0.79	−0.16	0.04	−0.31	−0.02	−0.30
Fuel	1.67	−0.28	1.37	0.10	1.00	0.00	0.86	−0.08
Drink and Tobacco	1.22	−0.60	1.22	−0.62	1.37	−0.67	1.36	−0.68
Transport and Communication	1.23	−1.21	1.73	−0.92	1.14	−1.23	1.05	−1.17
Other goods	1.21	−0.72	1.15	−0.77	2.03	−0.52	1.92	−0.47
Other services	1.40	−0.93	1.28	−0.78	1.03	−0.78	1.06	−0.74

is rejected, the *D.W.* statistic shows a sharp fall in each case.

The failure of homogeneity is not a new result (see, for example, Barten; Ray Byron; Deaton, 1974a), and can be ascribed to a number of possible causes. However, as far as we are aware, the introduction of serial correlation through the imposition of homogeneity is a result which has not been previously remarked, although it may have been implicit in earlier work. There are a number of plausible explanations for this phenomenon. For example, expenditure on several items may be relatively inflexible in the short run; housing is the obvious case here. The explanation of such items may require other variables such as stocks, lagged dependent variables, or time trends which can perhaps be proxied by the absolute price level. The omission of such variables will thus lead to a rejection of homogeneity associated with an introduction of serial correlation in the residuals of the restricted equations. In principle one could easily include such conditioning variables in the *AIDS* cost function, for example by allowing the αs to vary linearly with them, and this is likely to be an important topic for future research. A second explanation is the omission of price expectations—the argument advanced by Deaton (1977a) would suggest that factors such as the frequency of purchase for different goods will be relevant in assessing the response of expenditures to changes in price, especially when there is rapid relative or absolute price change. A third possibility, suggested by the discussion of aggregation in Section I, Part A above, is that it may be incorrect to assume that k, the index reflecting the distribution of household budgets and demographic structure, is independent of the average budget and the price vector. Finally, the assumption of weak intertemporal separability of nondurable goods in the intertemporal utility function, which is required to justify the conventional static utility-maximizing model, may be inappropriate. It is not difficult to construct other models which produce the result and without extensive further empirical work it is extremely difficult to discriminate between them.

In moving to the symmetric estimates (not reported here), in which P^* is replaced by P, we must first check the closeness of the approximation. Table 3 reproduces the two series, $P^*=\exp\{\Sigma w_k \log p_k\}$ and P scaled to be unity in the base year, i.e., $\exp\{\Sigma \alpha_k \log p_k + \frac{1}{2}\Sigma\Sigma \gamma_{kj} \log p_k \log p_j\}$ evaluated at the symmetric parameter estimates. Both series are based on unity in 1970. Clearly the differences are small; the absolute magnitude of the difference is never greater than .008.

The reestimation of the unconstrained and homogeneous models using P rather than P^* confirmed the empirical unimportance of the difference. Both sets of likelihoods are given in Table 4. It must be reemphasized that these findings are conditional on the kind of relative price movements that took place in our sample. However, even if relative price changes had been greater, the results suggest that the proce-

TABLE 3—COMPARISON OF PRICE INDICES

	P^*	P		P^*	P		P^*	P
1954	0.566	0.571	1961	0.684	0.686	1968	0.894	0.888
1955	0.587	0.595	1962	0.712	0.715	1969	0.946	0.944
1956	0.611	0.617	1963	0.729	0.730	1970	1.000	1.000
1957	0.631	0.636	1964	0.754	0.758	1971	1.084	1.084
1958	0.648	0.653	1965	0.793	0.797	1972	1.161	1.161
1959	0.655	0.661	1966	0.827	0.830	1973	1.271	1.279
1960	0.663	0.666	1967	0.851	0.852	1974	1.465	1.461

TABLE 4—COMPARATIVE VALUES OF 2 *LOG* LIKELIHOOD

	Levels		First Differences	
	Using P^*	Using P	Using P^*	Using P
Unrestricted	1722.5	1723.8	1560.0	1560.3
Homogeneous	1579.6	1585.1(7)	1546.6	1547.9(7)
Symmetric	–	1491.0(21)	–	1508.8(21)

Note: Number of restrictions in parentheses. Numbers can only be compared within columns, not between levels and first differences.

dure of starting with P^*, calculating *OLS* regressions, computing a new P, and repeating will be a computationally efficient way of obtaining good estimates of the full nonlinear system.

Symmetry, unlike homogeneity, cannot be tested on an equation-by-equation basis and we must rely on a large-sample likelihood-ratio test for the system as a whole. For comparison, twice the logarithm of the likelihood is 1722.5 for the unconstrained system, 1579.6 for the homogeneous model (a fall which reflects the individual restrictions) and this falls further to 1491.0 under symmetry. Since symmetry embodies twenty-one constraints over and above the seven of homogeneity, the restriction is rejected on an asymptotically valid χ^2-test whether or not the maintained hypothesis is taken to contain homogeneity. Once again, this is consistent with earlier results although rejection of symmetry *given* homogeneity is not always clear-cut in the studies cited above. The interpretation of the rejection is not clear without some convincing explanation of the lack of homogeneity. Without this, it is impossible to know whether or not we should expect symmetry to hold. For example, it is possible to introduce habits into demand functions so that, if they are allowed for, symmetry can be expected to hold, while if ignored, symmetry will be destroyed.

The full set of symmetric parameter estimates are not included here for reasons of space. The most interesting property of these, apart from symmetry, is in their implications for negativity. To assess this, the K matrix of equation (14) was evaluated for each year in the sample and its eigenvalues calculated. One of these is identically zero and, for concavity, the others should be negative. Contrary to this, we found one positive eigenvalue for the early part of the period, increasing to two by the end. The most obvious symptom of nonconcavity in the symmetric estimates was an estimated positive compensated own-price elasticity for fuel throughout the sample period. This may seem to be of limited importance given that the symmetric homogeneous model has already been rejected; if the cost function doesn't exist, why worry about its concavity? However, for several reasons it would be extremely useful to have parameter estimates for a reasonably general concave homogeneous cost function. For example, we frequently wish to calculate price and quantity index numbers or to use optimal taxation formulae to derive numerical values for tax rates. All such calculations require numerical estimates of cost functions and, if they are to make any sense at all, these cost functions must be both homogeneous and concave. Consequently, in cases where empirical estimates of demand equations have been used in applied welfare analysis, the linear expenditure system has invariably been used; see, for example, Anthony Atkinson and Joseph Stiglitz, Muellbauer (1974), or Deaton (1977b). With the linear expenditure system, the model is so restrictive that concavity of the cost function is virtually guaranteed provided inferior goods do not appear. But this restrictiveness is also known to be empirically false (see, for example, Deaton, 1974b or 1978) so that it would be of considerable value to have estimates of a concave cost function which allowed considerably more substitution than does the linear expenditure system. Consequently, it will be of considerable interest in future work to attempt to restrict the parameters further so that the estimated cost function is concave.

Finally, we turn to the estimation of the model in first-difference form. Here we use equation (17) plus intercepts, i.e.,

$$(21)\quad \Delta w_i = \eta_i + \beta_i \Delta log\left(\frac{x}{P}\right) + \sum_k \gamma_{ik} \Delta log p_k$$

where the constants η_i are introduced primarily for econometric reasons but, if significant, would imply time trends in the original model which expresses the variables in levels. The P is taken as in Table 3. In these regressions homogeneity is only rejected for food and for transport and communication; clothing and housing, which rejected homogeneity in the earlier regressions, now yield insignificant F-ratios. Closer inspection reveals that for both these cases, the constant term η_i, which is insignificant without constraints, becomes significant when homogeneity is imposed. Similarly, for transport and communication η_i becomes significant when homogeneity is imposed, but in this case the F-ratio remains significant. This would support our earlier conjectures as to the possible role of time trends, stocks, or other omitted variables in explaining nonhomogeneity. Likewise, in Table 2 the expenditure elasticities from the first-difference model tend to be higher than the levels estimates when stock effects are likely to be important (clothing, other goods) and lower when one would expect short-run total expenditure effects to be limited (food, housing, transport). Otherwise, the first-difference parameter estimates, homogeneous or unconstrained, are rather close to the values originally obtained. As with levels, tests of concavity with the first-difference model revealed several violations. The likelihoods for the two models are summarized in Table 4; the fact that homogeneity cannot be rejected overall at the 5 percent level reflects the importance of the time trends in the housing, clothing, and transport and communication equations. Note too that in this case, from the last column of the table, symmetry is only just rejected given homogeneity. Hence, if we make some allowance for the asymptotic nature of the test, these final results would suggest that the introduction of (arbitrary) time trends removes much of the conflict between the data and the hypothesis of a representative consumer maximizing a conventional static utility function.

III. Summary and Conclusions

In this paper we have introduced a new system of demand equations, the *AIDS*, in which the budget shares of the various commodities are linearly related to the logarithm of real total expenditure and the logarithms of relative prices. The model is shown to possess most of the properties usually thought desirable in conventional demand analysis, and to do so in a way not matched by any single competing system. Fitted to postwar British data, the *AIDS* is capable of explaining a high proportion of the variance of the commodity budget shares but, unless allowance is made for omitted variables by the arbitrary use of time trends, does so in a way which is inconsistent with the hypothesis of consumers making decisions according to the model's demand functions governed by the conventional static budget constraint. These results suggest that influences other than current prices and current total expenditure must be systematically modelled if even the broad pattern of demand is to be explained in a theoretically coherent and empirically robust way. Whether these developments generalize the static framework by including stock effects, errors in price perceptions, or by going beyond the assumption of weak intertemporal separability on which the static model rests, we believe that the *AIDS*, with its simplicity of structure, generality, and conformity with the theory, offers a platform on which such developments can proceed.

APPENDIX: *AIDS* IN THE CONTEXT OF AGGREGATION THEORY

In Muellbauer (1975, 1976), a definition of the existence conditions for a representative consumer is given which allows more general behavior than the parallel linear Engel curves which are required if average demands are to be a function of the average budget. We know that in general, the average budget share

$$\bar{w}_i = \sum_h p_i q_{ih} \Big/ \sum_h x_h \equiv \sum_h x_h w_{ih} \Big/ \sum_h x_h$$

is a function of prices and the complete distribution vector $(x_1, x_2, \ldots, x_H)$. A representative consumer exists in Muellbauer's sense if each $\bar{w}_i$ can be written as a function of prices and the same single scalar x_0, itself a function of prices and the distribution vector. This scalar, which can be thought of as marking a position in the distribution of *x*s, is the representative budget level. Muellbauer shows that for an x_0 to exist such that

(A1)

$$\sum_h x_h w_{ih}(x_h, p) / \Sigma x_h = w_i\{x_0(x_1, \ldots, x_H, p), p\}$$

the individual budget share equations must have the "generalized linear" (*GL*) form:

(A2)

$$w_{ih}(x_h, p) = v_h(x_h, p) A_i(p) + B_i(p) + C_{ih}(p)$$

where v_h, A_i, B_i, and C_{ih} are functions satisfying $\Sigma_i A_i = \Sigma_i C_{ih} = \Sigma_h C_{ih} = 0$ and $\Sigma B_i = 1$. Clearly, (A1) goes beyond the usual formulation of $x_0 = \bar{x}$, and, as we shall see below, allows us to incorporate into the demand functions features of the expenditure distribution other than the mean.

Of particular interest is the case where x_0 is independent of prices, depending only on the individual *x*s. This occurs if, and only if, the v_h function in (A2) restricts to

$$\text{(A3)} \quad v_h(x_h, p) = \{1 - (x_h / k_h)^{-\alpha}\}^{\alpha^{-1}}$$

where α is a constant and k_h, although not a function of x_h, and p is free to vary from household to household. In this case, the budget shares are said to have the "price-independent generalized linear" form (*PIGL*). Note the special case of (A3) as $\alpha \rightarrow 0$, i.e.,

$$\text{(A4)} \qquad v_h(x_h, p) = log(x_h / k_h)$$

For obvious reasons, this is referred to as the *PIGLOG* case. By substituting (A3) in (A2), (A1) can be used to give an explicit form for x_0, viz.,

$$\text{(A5)} \quad x_0 = \left\{ \Sigma \left(\frac{x_h}{k_h} \right)^{-\alpha} \Big/ \Sigma x_h \right\}^{-1/\alpha}$$

If we assume that individual behavior is preference consistent, the cost function corresponding to *PIGL* takes the form

$$\text{(A6)} \quad \{c(u_h, p) / k_h\}^{\alpha} = (1 - u_h)\{a(p)\}^{\alpha} + u_h\{b(p)^{\alpha}\}$$

which as α tends to zero takes the *PIGLOG* form

$$\text{(A7)} \quad log\{c(u_h, p) / k_h\} = (1 - u_h) log\{a(p)\} + u_h log\{b(p)\}$$

where $a(p)$ and $b(p)$ are linear homogeneous concave functions, α is the constant parameter of (A3), and (with some exceptions discussed below) $0 \leqslant u \leqslant 1$. The quantity k_h can be used to allow for family composition effects within *PIGL*; for the standard or "reference" household k_h is unity.

Since the *AIDS* is a member of the *PIGLOG* family, and hence of *PIGL*, we can achieve maximum generality by discussing some of the important properties of this class. If, omitting the household subscript, we write q_i for the quantity demanded of good i, then, by the derivative property of the cost function, $q_i = \partial c/\partial p_i$ so that $w_i = p_i q_i / x = \partial \log c / \partial \log p_i$. Hence from (A6), taking $k_h = 1$, and differentiating

$$\alpha c^{\alpha} \frac{\partial \log c}{\partial \log p_i} = \alpha a^{\alpha} a_i (1-u) + u \alpha b^{\alpha} b_i \tag{A8}$$

where $a_i = \partial \log a / \partial \log p_i$ and $b_i = \partial \log b / \partial \log p_i$. Hence, substituting x for c,

$$w_i = (1-u)\left(\frac{a}{x}\right)^{\alpha} a_i + u\left(\frac{b}{x}\right)^{\alpha} b_i \tag{A9}$$

where, from (A6), $u = (x^{\alpha} - a^{\alpha})/(b^{\alpha} - a^{\alpha})$ or from (A7), $u = (\log x - \log a)/(\log b - \log a)$. Similarly, when $\alpha = 0$,

$$w_i = (1-u) a_i + u b_i \tag{A10}$$

Equations (A9) and (A10) have attractive interpretations. Cost $c(u,p)$ is increasing in utility as long as $b(p)$ is greater than $a(p)$—note that this does not depend on the sign of α—so that as u increases from 0 to 1, $c(u,p)$ increases from $a(p)$ to $b(p)$ with w_i moving from a_i to b_i. Hence a total expenditure of $a(p)$ can be thought of as "poverty" expenditure with associated expenditure pattern a_i, while $b(p)$ is "affluence" expenditure with budget shares b_i. On this interpretation $x = a(p)$ and $x = b(p)$ are the equations of the tangents to the poverty and affluence indifference curves, respectively, $u = 0$ and $u = 1$. From (A6) therefore, we see that the tangent to the indifference curve actually attained is the mean of order α of poverty and affluence tangents, the weights depending on the welfare level or outlay of the household. This averaging is even more obvious in the value share equations (A9) and (A10). Since $(1-u)(a/x)^{\alpha}$ and $u(b/x)^{\alpha}$ sum to unity, as do $(1-u)$ and u, these equations give the actual budget shares as weighted averages of a_i and b_i.

Since the value shares of luxuries increase with total outlay and hence with u, we can characterize luxuries and necessities simply by whether b_i is greater than or less than a_i. Inferior goods are not excluded under *PIGL* and it is straightforward to construct examples from both (A9) and (A10).

Note finally that there are restrictions on the possible set of x and p over which the cost function and the associated demands are valid. One set of restrictions is implied by the necessity that, for all $i, 0 \leqslant w_i \leqslant 1$. The upper bound is relevant when $b_i > a_i$ and implies that $u = (x^{\alpha} - a^{\alpha})/(b^{\alpha} - a^{\alpha}) \leqslant \min_i[\{1 - a_i(a/x)^{\alpha}\}/\{(b/x)^{\alpha} b_i - (a/x)^{\alpha} a_i\}]$. The lower bound, relevant when $b_i < a_i$ requires similarly that $u = (x^{\alpha} - a^{\alpha})/(b^{\alpha} - a^{\alpha}) \geqslant \max_i[(a/x)^{\alpha} a_i / \{(a/x)^{\alpha} a_i - (b/x)^{\alpha} b_i\}]$. The second set of restrictions are those required to ensure that the cost function is concave. From (A6), we can see that a *sufficient* condition for concavity of $c(u,p)$ is that $a(p)$ and $b(p)$ be concave and that $0 \leqslant u \leqslant 1$. However, this is by no means necessary. If $b(p)$ is "more concave" than $a(p)$, then $c(u,p)$ is concave for $u > 0$ and for a range of $u > 1$. It can be shown that the *PIGL* cost function is concave for *all* $x > 0$, $p > 0$ if and only if $a_i = b_i$ for all i, and $a(p)$ and $b(p)$ are concave. In this not very interesting case, preferences are homothetic.

The practical application of the *PIGL* class requires selection of specific functional forms for the functions $a(p)$ and $b(p)$; those leading to the *AIDS* have been discussed in the text. However, the *PIGL* class is related to two other well-known models. Note first that if $\alpha = 1$, and $k_h = 1$, (A6) becomes the Gorman polar form. The *PIGL* class thus includes all models with linear Engel curves, for example, linear expenditure system, the quadratic utility function, as special cases. Perhaps less obviously, a weakly restricted form of the indirect translog is also *PIGLOG*. From Jorgenson and Lau the translog indirect utility function is

$$\text{(A11)}\quad u = a_0 + \sum_i \alpha_i \log\left(\frac{p_i}{x}\right) + \frac{1}{2}\sum_i\sum_j \beta_{ij} \log\left(\frac{p_i}{x}\right)\log\left(\frac{p_j}{x}\right)$$

where we can choose $\Sigma\alpha_i = -1$ and $\beta_{ij} = \beta_{ji}$ as arbitrary normalizations. Write $\Sigma_k \beta_{ki} = \beta_{Mi}$, then if we impose the additional restriction that $\Sigma_i \beta_{Mi} = 0$, (A11) solves explicity for $\log c(u,p)$ to

$$\text{(A12)}\quad \log c(u,p) = \frac{u - \frac{1}{2}\Sigma\Sigma\, \beta_{ij} \log p_i \log p_j - \Sigma\alpha_i \log p_i - a_0}{1 + \sum_i \beta_{Mi} \log p_i}$$

This is of the general form $\log c(u,p) = \log a(p) + u/\log h(p)$ for appropriate choice of $a(p)$ and $h(p)$. Using $\log h(p) = 1/\log\{b(p)/a(p)\}$ to define $b(p)$ and substituting, we see that (A12) is identical to (A7). Hence, in this case $(\Sigma_i \beta_{Mi} = 0)$ and in this case only, the indirect translog allows consistent aggregation. No such result holds for any interesting subcase of the direct translog.

REFERENCES

A. B. Atkinson and J. E. Stiglitz, "The Structure of Indirect Taxation and Economic Efficiency," *J. Publ. Econ.*, Apr. 1972, *1*, 97–119.

A. P. Barten, "Maximum Likelihood Estimation of a Complete System of Demand Equations," *Euro. Econ. Rev.*, Fall 1969, *1*, 7–73.

E. R. Berndt, M. N. Darrough, and W. E. Diewert, "Flexible Functional Forms and Expenditure Distributions: An Application to Canadian Consumer Demand Functions," *Int. Econ. Rev.*, Oct. 1977, *18*, 651–75.

R. P. Byron, "A Simple Method for Estimating Demand Systems under Separable Utility Assumptions," *Rev. Econ. Stud.*, Apr. 1970, *37*, 261–74.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Transcendental Logarithmic Utility Functions," *Amer. Econ. Rev.*, June 1975, *65*, 367–83.

A. S. Deaton, (1974a) "The Analysis of Consumer Demand in the United Kingdom, 1900–1970," *Econometrica*, Mar. 1974, *42*, 351–67.

______, (1974b) "A Reconsideration of the Empirical Implications of Additive Preferences," *Econ. J.*, June 1974, *84*, 338–48.

______, *Models and Projections of Demand in Post-War Britain*, London 1975.

______, (1977a) "Involuntary Saving through Unanticipated Inflation," *Amer. Econ. Rev.*, Dec. 1977, *67*, 899–910.

______, (1977b) "Equity, Efficiency and the Structure of Indirect Taxation," *J. Publ. Econ.*, Apr. 1977, *8*, 299–312.

______, "Specification and Testing in Applied Demand Analysis," *Econ. J.*, Sept. 1978, *88*, 524–36.

W. E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *J. Polit. Econ.*, May/June 1971, *79*, 481–507.

______, "Applications of Duality Theory," in Michael D. Intriligator and David A. Kendrick, eds, *Frontiers of Quantitative Economics*, Vol. 2, Amsterdam 1974, ch. 3.

D. W. Jorgenson and L. J. Lau, "The Structure of Consumer Preferences," *Annals Econ. Soc. Measure.*, Winter 1975, *4*, 49–101.

C. E. V. Leser, "Forms of Engel Functions," *Econometrica*, Oct. 1963, *31*, 694–703.

______, "Income, Household Size and Price Changes 1953–1973," *Oxford Bull. Econ. Statist.*, Feb. 1976, *38*, 1–10.

G. E. Mizon, "Inferential Procedures in Non-Linear Models: An Application in a U.K. Industrial Cross-Section Study of Factor Substitution and Returns to Scale," *Econometrica*, July 1977, *45*, 1221–42.

J. Muellbauer, "Recent U.K. Experience of Prices and Inequality: An Application of True Cost of Living and Real Income Indices," *Econ. J.*, Mar. 1974, *84*, 32–55.

______, "Aggregation, Income Distribution and Consumer Demand," *Rev. Econ.*

Stud., Oct. 1975, *62*, 525–43.

______, "Community Preferences and the Representative Consumer," *Econometrica*, Sept. 1976, *44*, 979–99.

R. W. Shephard, *Cost and Production Functions*, Princeton 1953.

______, *Theory of Cost and Production Functions*, Princeton 1970.

J. R. N. Stone, *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom, 1920–1938*, Vol. 1, Cambridge 1953.

______, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Econ. J.*, Sept. 1954, *64*, 511–27.

Henri Theil, "The Information Approach to Demand Analysis," *Econometrica*, Jan. 1965, *33*, 67–87.

______, *Statistical Decomposition Analysis with Applications in the Social and Administrative Sciences*, Amsterdam 1972.

______, *Theory and Measurement of Consumer Demand*, Vols. 1 and 2, Amsterdam 1976.

H. Working, "Statistical Laws of Family Expenditure," *J. Amer. Statist. Assn.*, Mar. 1943, *38*, 43–56.

A Structural Model of Murder Behavior and the Criminal Justice System

By STEPHEN A. HOENACK AND WILLIAM C. WEILER*

A recent article by Isaac Ehrlich (1975a) in this *Review* presented support for the hypothesis that capital punishment deters individuals from committing murder. Because of the elegant theory of deterrence underlying it and the close correspondence of the empirical estimates to the theoretical hypotheses, the article has been widely influential. Since the public appears to take deterrence into account in determining the appropriateness of execution as a punishment for murder, the correctness of this study is of major significance. Perhaps for this reason, a large body of literature critical of Ehrlich's study has been developed.[1] However, little attention has been given to the identification of his "murder supply function," an equation representing the response of murder behavior to the punishments meted out by society.[2] By providing estimates of a murder supply function in the context of a structural model in which the identification of each equation is examined, we suggest that his estimates of this function may actually represent the response of society's punishments to murder behavior.[3]

Identification is a property of a model based on the theory underlying it and is thus logically prior to estimation. However, Ehrlich appears to have treated identification as a property of his estimation technique rather than of his model. In describing an instrumental variable technique for estimating the murder supply function, he listed the variables against which he regressed each of the penalties taken as endogenous.[4] The calculated values of the endogenous penalties from these regressions were substituted in the estimation of the murder supply function. This procedure was not based on specified equations representing the response of society to homicides in the determination of penalties.[5] Without such equations it was not possible to evaluate the identification of the murder supply function based on the order condition, nor could the reduced-form equations have been specified correctly.

---

*Professor, Humphrey Institute of Public Affairs and director, Management Information Division; and associate director, Management Information Division, University of Minnesota, respectively. We gratefully acknowledge the comments of Colleen T. Davidson, Gestur B. Davidson, Stephen E. Fienberg, Edward M. Foster, Robert T. Kudrle, William S. Reece, Christopher S. Sims, Walter Vandaele, an anonymous referee, and the managing editor on earlier drafts of this paper. Jeffrey A. Rivkin, Sharleen Goldberg, and Betsy A. Flaten provided valuable suggestions and research assistance.

[1] The following list of critiques of Ehrlich's work is incomplete as this literature is growing rapidly. See David Baldus and James Cole; William Bowers and Glenn Pierce; Phillip Cook; Lee Friedman; Peter Passell and John Taylor. Published replies to Ehrlich's critics appear in Ehrlich (1975b, 1977a), and Ehrlich and Joel Gibbons.

[2] Brian Forst, Victor Filatov, and Lawrence R. Klein, and Franklin Fisher and Daniel Nagin have discussed the identification of Ehrlich's murder supply function, although they did not actually specify estimable equations for the criminal justice system or for society's provision of resources to it. Comments on these analyses are found in Ehrlich and Randall Mark (1977).

[3] This study is based on a time-series sample and its results are comparable to those of Ehrlich's time-series study (1975a). Ehrlich (1977b) has also performed a cross-section study which does not contain a full treatment of identification. We believe that it would be difficult to estimate an identified structural cross-section model with currently available data.

[4] See Ehrlich's note to Table 3 (1975a, p. 410). Identification seems to be treated as a property of estimation in Ehrlich (1977b), in which he explains his inability to provide estimable equations other than the crime supply function as follows: "Due to data exigencies, the parameters of the supply functions are estimated via classical least squares and seemingly unrelated simultaneous equation techniques rather than the two-stage least squares procedures employed in my related works" (p. 749).

[5] Other authors have explicitly specified equations for the response of society to nonmurder criminal behavior, although empirical tests for the correctness of the overidentifying restrictions for each equation were not performed. These include cross-section analyses of state crime rates (Ehrlich, 1973), of crime rates in California counties (Llad Phillips and Harold Votey), and of crime rates by census tract in Rochester, New York (Richard Thaler).

The specification of estimable equations of the response of society to homicide behavior requires a theory of this response. In a section titled "Defense Against Murder" (1975a), Ehrlich discussed in general terms the rational behavior of enforcement agencies based on a posited social loss function. Yet this discussion did not provide a basis for explicitly specifying equations representing the endogenous behavior of enforcement agencies that contributed to the causality underlying observed data on penalties and murders. Without leading to these equations, his theoretical discussion of such issues as the social damage resulting from the deaths of victims, the costs of dealing with and punishing offenders, the optimal rate of executions, and tradeoffs between conviction and execution rates, does not contribute to the identification of his murder supply function.

Because the identification of a murder supply function requires an exact knowledge of each equation in the complete structural model, the present study provides a structural model, originated by Hoenack, Robert Kudrle, and David Sjoquist (hereafter, H-K-S), that includes equations representing the response of society to murder behavior. While the murder supply function in this model is identified according to the order condition, the empirical hypothesis of the correctness of the zero restrictions on the coefficients of the predetermined variables excluded from it must be rejected. When respecified in a way which is justifiable on a priori grounds, and which leads the equation to pass the empirical test, the murder supply function no longer supports the hypotheses that execution and conviction risks deter murderers. We present empirical support for an alternative interpretation that the response of society to murder behavior, not murder behavior itself, causally generated the apparent support for these hypotheses in Ehrlich's murder supply function.

Section I describes the structural model and discusses the a priori identification of its equations. Section II presents estimates of the model's coefficients and the results of empirical tests of the validity of the zero restrictions for the coefficients of the excluded predetermined variables in each equation. Section III provides the alternative interpretation of the causality underlying Ehrlich's estimates.

I. The Structural Model

A. *Overview*

Equations (1)–(9) present the structural model representing the behavior of murderers, the criminal justice system, and society in causally generating observed time-series data covering the years 1933–69 on homicides and penalties. (The variables are defined in Table 1.) Equations (7)–(9) are identities. Equation (1) is a murder supply function specified with the same variables employed by Ehrlich. Equations (2)–(6) are based on the H-K-S model. The model's three production equations (2)–(4) reflect the criminal justice activities of apprehending suspects, convicting them, and upholding appeals of death sentences. Murders, apprehensions, and convictions represent workloads in these production equations. Resources that society provides to the criminal justice system also appear in these equations. The level of these resources is determined in equations (5) and (6). These equations are public expenditure functions which relate variables that stand as proxies for attitudes about the social loss from murders and other crimes to the collective funding of criminal justice activities.

Within equations (2)–(4), increases in murders augment workloads in the criminal justice system which, in the absence of additional resources, result in lower levels of the measured rates of apprehension of suspects, conviction of apprehended suspects, and execution of convicted criminals. Because of the high costs of renegotiating budgets, H-K-S hypothesized that society will change the amount of resources allocated to the criminal justice system only if a change in the number of murders is perceived as permanent. Thus, as specified in equations (5) and (6), resources can be expected to have a lagged response to workloads.

$$(1)\quad \frac{Q}{N}=a_{11}+a_{12}\log Pa + a_{13}\log Pc|a+a_{14}\log Pe|c+a_{15}\log A_1 + a_{16}\log L+a_{17}\log Y_p+a_{18}\log U+a_{19}\mathrm{T}$$

$$(2)\quad \log Pa=a_{21}+a_{22}\log Q + a_{23}\log POL+a_{24}\log NW$$

$$(3)\quad \log Pc|a=a_{31}+a_{32}\log A + a_{33}\log R+a_{34}\log NW$$

$$(4)\quad \log Pe|c=a_{41}+a_{42}\log C + a_{43}\log R+a_{44}CL + a_{45}\mathrm{T}\times \log C_{-1}+a_{46}\log A_4$$

$$(5)\quad \log R=a_{51}+a_{52}\log Q+a_{53}\log Y_p + a_{54}\log N+a_{55}\log A_2+a_{56}\log NW + a_{57}\log A_3+a_{58}\log Q_{-1}+a_{59}\log Q_{-2} + a_{5,10}\log OC_{-1}+a_{5,11}\log OC_{-2}$$

$$(6)\quad \log POL=a_{61}+a_{62}\log Q+a_{63}\log Y_p + a_{64}\log N+a_{65}\log A_2 + a_{66}\log NW+a_{67}\log A_3+a_{68}\log Q_{-1} + a_{69}\log Q_{-2}+a_{6,10}\log OC_{-1} + a_{6,11}\log OC_{-2}$$

$$(7)\quad Q\equiv N(Q/N)$$

$$(8)\quad A\equiv QPa$$

$$(9)\quad C\equiv QPaPc|a=A\,Pc|a$$

TABLE 1—DEFINITIONS OF VARIABLES IN THE MODEL

Endogenous Variables
Q = observed murders (capital murders plus other murders less negligent homicides)
A = arrests for observed murders
C = convictions for observed murders
Pa = the probability of apprehension
$Pc|a$ = the conditional probability of conviction given apprehension
$Pe|c$ = the conditional probability of execution given conviction
R = total resources available to the criminal justice system which may be allocated to the arresting, convicting, and executing of capital murderers
POL = resources available to police only
Exogenous Variables
L = proportion of the population in the labor force
U = percent of the labor force unemployed
Y_p = Milton Friedman's estimate of (real) permanent income per capita
A_1 = proportion of the population in the 14–24 age group
T = chronological time
N = civilian population (in thousands)
NW = proportion of the population which is nonwhite
A_2 = proportion of the population within the 25–49 age group
CL = dummy variable representing class action appeals and taking the value 1 for 1965–69
A_3 = proportion of the population age 50 and above
OC = other violent crimes
A_4 = proportion of the population within the 14–19 age group

Note: Data used are very similar to those used by Ehrlich. Our modifications of Ehrlich's data are described in the Appendix. A detailed listing of our data sources and the actual data are available on request.

From Ehrlich's "Defense Against Murder" (1975a) and from our structural model, it can be seen that both analyses of society's response to murder behavior assume rational behavior on the part of society in allocating resources to the criminal justice system, but they differ in the one respect crucial for identification. The structural model provides equations that could have jointly caused, along with a murder supply function, the data used to estimate that function. Ehrlich's analysis does not provide such equations.

B. *Description and Theoretical Justification of the Equations in the Structural Model*

1. *The Murder Supply Function*

The murder supply function was initially specified in the structural model with the same variables that Ehrlich employed. Other authors have found that Ehrlich's empirical support for his theory's prediction of ascending negative coefficients for the deterrence variables does not hold under many alternative specifications.[6] We chose to examine the identification of a murder supply function having the same results as Ehrlich's in terms of statistically significant negative ascending coefficients for the deterrence variables when estimated outside a structural model. We measured Pa (apprehension risk) and $Pc|a$ (conditional conviction risk) in the same way Ehrlich measured them. Among the alternative measures of execution risk used by Ehrlich, we used his $PXQ1_{-1}$, the ratio of current period executions to the previous year's convictions.[7] We used a semilog formulation for the murder supply function rather than the *log*-linear form used by Ehrlich.[8]

2. *The Production Equations*

The H-K-S theory holds that adding resources to the criminal justice system (POL in equation (2) and R in equations (3) and (4)) increases the apprehension, conviction, and execution rates. This hypothesis is based on the assumption that police and prosecutors invariably lack resources to maximize the numbers of apprehensions, convictions, and upheld convictions with death sentences, and that at least some of any additional resources are devoted to increasing these outputs. The theory hypothesizes that increases in the workload variables, Q in equation (2), A in equation (3), and C in equation (4), respectively, reduce the apprehension, conviction, and execution rates. This prediction depends on the assumption that a larger workload cannot be handled as successfully as a smaller workload with a given amount of resources. However, police and prosecutors can redeploy resources from more- to less-resource-absorbing cases to mitigate the reduced production brought about by increased workloads. Based on an institutional analysis of the criminal justice system, the H-K-S theory predicts that police are more able than prosecutors to redeploy resources among cases. Thus the predicted negative coefficient for murders for the production of apprehensions in equation (2) is larger than the negative coefficients for apprehensions and convictions in the production of convictions and executions in equations (3) and (4). An increase in the measured workload in the execution equation includes actual workload unrelated to upholding appeals of death sentences because the data for convictions include all convictions for murders.

---

[6] Ehrlich's theory provides justification for the deterrence variables in his murder supply function including a priori prediction of the ascending values of their coefficients. However, he provides less definite justification for the other variables in this equation.

[7] Pa and $Pc|a$ are measured by the clearance rate for murders and nonnegligent manslaughters and the percent of individuals found guilty as charged for murder and nonnegligent manslaughter, respectively. Ehrlich uses three basic measures of execution risk: $PXQ2$ (the ratio of the current year's executions to convictions), $PXQ1$ (the ratio of next year's executions to current year convictions), and $PXQ1_{-1}$ (the ratio of current executions to previous year's convictions). While we have no strong preference, we use $PXQ1_{-1}$ because it allows for a lag between conviction and execution and does not require potential offenders to forecast executions.

[8] Ehrlich specified a *log*-linear relation between Q^c/N (the capital murder rate) and the right-hand side variables in the murder supply function. Because of a lack of data on Q^c, he then specified a time trend relation between Q (total murders for which data are available) and Q^c to transform the dependent variable to Q and in which time was required to enter the murder supply function as T itself. When *log*-linear forms are specified for any two of Q^c/N, Q^n/N, and Q/N (where Q^n is the number of noncapital murders), the linear identity $Q/N \equiv Q^c/N + Q^n/N$ implies an unusual intrinsically non-linear form for the third ratio. An alternative specification of a semilog formulation for Q^n/N and Q^c/N yields a semilog formulation for Q/N whose parameters can be estimated using linear regression. Thus the semilog relationship is consistent with a straightforward relationship between Q and Q^c endogenously determined within the theory of murder behavior. Time is included simply as a separate independent variable to control for omitted exogenous influences on Q^c/N and/or Q^n/N; thus while the form of the variable (i.e., T or *ln*T or T^2, etc.) influences the empirical results, its a priori specification is clearly arbitrary.

For this reason, H-K-S argue that the negative coefficient for apprehensions in equation (3) will be larger than the negative coefficient for convictions in equation (4).

In the production equation for executions we included the following variables to control for the rapid growth of appeals over the 1933–69 sample period. A dummy variable was entered for the class action suits, initiated in 1965 by Anthony Amsterdam and his colleagues, that contributed to a near cessation of executions after 1964 and a cessation by 1968.[9] We also incorporated an interaction term between chronological time and lagged convictions to control for the steady increase in the lag from sentencing to execution over the sample period.[10]

To control for possible differences in the production relations between urban and rural areas, we wished to include in them the fraction of the population in urban areas as a variable. Because available data on urban populations are not sufficient for estimation, we used the proportion of the population that is nonwhite (NW)[11] as a crude proxy variable in the production equations for apprehensions and convictions. We found that NW would not enter the production equation for executions when the interaction term in equation (4) was also entered.[12] This result may be due to the simple correlation coefficient of 0.95 between the *log* of NW and the time trend. We retained the interaction term in the production equation for executions and excluded NW. Thus the net influence of NW (see fn. 11) is embedded in other estimated coefficients.

In the production equation for executions we entered the fraction of the population in the 14–19 age group to control for the reluctance of society to execute youthful offenders.[13]

3. *Expenditure Equations*

The variables in our expenditure functions measure public concern and willingness to expend resources on the criminal justice system.[14] They do not necessarily imply that the public is informed about the effects on criminal behavior of spending varying levels of resources on criminal justice activities.[15] We included real income because this variable could influence the willingness of the public to spend resources on criminal justice activities. For example, one might specify the values that society is willing to pay to avoid the loss of human life as the cost of murder, which in turn could be a function of real income.[16]

---

[9] Two of the decisions argued by the Amsterdam organization were *Adderly v. Wainwright* and *Hill, Saterfield, and Anderson v. Nelson.*

[10] While data on the time delays were not systematically collected until 1960, available estimates suggest that during the 1930's execution usually followed conviction within a year (see Burton Wolfe, p. 82). The median length of time from sentencing to execution had increased to nearly four years by 1965 (see Hugo Bedau, p. 30). Increased delays resulted from a number of major court decisions which extended grounds for appeals. See, for example, *Norris v. Alabama; Brown et al. v. State of Mississippi; Mallory v. United States; Escobedo v. State of Illinois; People v. Morse; People v. Dorado.* With the exception of the Amsterdam suits, the effects of new grounds for appeal were more continuous in their influence than dummies would indicate. The impact of these other decisions developed gradually because of time delays in acquiring and learning how to apply information. Through educating other defendants, Caryl Chessman (1954; 1955; 1957) contributed to the widespread use of the appeals process.

[11] Alternative measures of *U.S.* urban populations are available only for census years. The percentage increases in the fraction of the resident (civilian not available) *U.S.* population living in Standard Metropolitan Statistical Areas were 2.1 for 1930–40, 4.9 for 1940–50, 7.0 for 1950–60 and 5.6 for 1960–70 (see Donald Bogue, Table 3-1, p. 47; and U.S. Bureau of the Census, No. 19, p. 20). The corresponding percent increases for our nonwhites variable (applicable to the civilian population) were 1.1, 4.9, 7.1, and 7.5. Aside from the possible role of race as a proxy for urbanization, the rates of apprehension, conviction, and execution may differ by race because of public opinion, and the weights accorded by police and prosecutors to crimes committed by different categories of the population and to the economic capability of defendants to provide for their defense. In William Landes' paper, a model of the behavior of prosecutors and defendants is provided that predicts plea bargains and trial outcomes given an index of the importance to the prosecutors of each crime (which can include the defendant's characteristics) and the resources available to the prosecutor and each defendant.

[12] Ehrlich found that when NW and a time trend were entered into the murder supply function, the coefficient of NW became statistically insignificant.

[13] We calculated the proportion of those arrested for murder in a number of different age groups who were subsequently executed. Those proportions are .012 for persons under 20, .018 for persons aged 20–24, .020 for persons aged 25–29, .018 for persons aged 30–34, .013 for persons aged 35–44, and .010 for persons over 44. While the proportions are lowest in the under-20 age group and in the over-35 age group, we entered only the fraction in the younger group, since the fraction of arrests in the over-35 age group could also capture the effect of the relatively more positive attitudes of persons in this age group towards capital punishment. This fraction did not enter significantly when entered with alternative lags perhaps because of changes in definitions for the arrest data over the sample period. For this reason we entered the fraction of the population in the 14–19 age group.

We entered the fractions of the population in the 25–49 and over-50 age groups to control for the possible effects of age on the demand for law enforcement. We were able to find only indirect evidence on this topic and the available evidence is based on sketchy data. There is some evidence that "socialization" or attitudes about social behavior may be influenced by events in the life cycle such as leaving school and beginning one's career, marrying and forming families, child rearing, purchasing a home, and retiring. (See Jeylan Mortimer and Roberta Simmons and references therein, and especially Orville Brim, Jr.) Most of those events occur within the former age group, but the older age group experiences the events of children leaving home, retirement planning, and retirement.[17] With references to attitudes about crime specifically, there is some evidence (see Monica Blumenthal et al., ch. 3) that fear of crime and the desire for larger police forces increase somewhat as age increases beyond the teenage years (see Hazel Erskine). We were, however, unable to find data tying socialization to voting behavior. We also included *NW* as a possible control for the influence of urbanization in the demand for law enforcement. Lagged as well as current murders were included because of the H-K-S hypothesis that adjustments of resources to changes in workloads are gradual. Similarly, lagged values of other violent crimes were entered into the equation.[18]

C. *A Priori Identification*

According to the counts of included jointly dependent variables and excluded predetermined variables, each equation is overidentified on a priori grounds. The exclusion from the murder supply function of variables controlling for appeal activity in the production relation for executions and of variables controlling for lagged murders and other violent crimes in the resource equations contributes to the identification of this equation on a priori grounds. Socioeconomic variables influencing murder behavior excluded from the production and resource relations identified these equations on a priori grounds. However, data limitations made it necessary to include only a crude proxy (*NW*) for the possibly important variable of the fraction of the population living in urban areas. Many of the model's predetermined variables are also highly intercorrelated over the time-series sample. (Examples of simple correlations are T and $logN$(0.98), $logA_3$ and $log\ Y_p$(0.95), $logNW$ and $logN$(0.93), and T and $log\ Y_p$(0.97).) Consequently, the measures used to control for the influence of variables in one equation can be controlling simultaneously for the influence of other variables in other equations. For this reason, even though the model's equations are overidentified according to counts of variables, the equations may not be estimable with available data.

---

[14] Similar variables were used by Jeffrey Chapman; Phillips and Votey; Thaler.

[15] Public expenditure functions (including our own) do not shed much light on the subject of how well informed the public is about the results of collective expenditures. In regard to public higher education expenditures, Charles Clotfelter found support for the hypothesis that the public is informed about some effects but appears to be uninformed about others.

[16] See E. J. Mishan.

[17] The mean age of marriage for men in the United States has varied between 23 and 26 in this century with the first child coming 1.5 to 3 years after marriage so that the 25–49 age range encompasses the years of family formation and child rearing. Home purchase is also concentrated in this age group (see Hugh Carter and Paul Glick).

[18] The model could include the contemporaneous rate of other violent crimes as an endogenous variable. However, estimation of other violent crimes was hampered by the lack of adequate data. There are no continuous time-series data on convictions for other violent crimes over the sample period; juvenile convictions are included from 1933–60 and excluded thereafter. Sentencing data are also unavailable for other violent crimes over the sample period. Unless additional predetermined variables are included in equations for the supply of other violent crimes and production equations associated with them, there would be no change in our reported results. These results would change, however, if a full-information estimation technique were used. Using three-stage least squares (*3SLS*) we estimated a crude model including equations for other violent crimes but lacking sentencing data and with an estimated series on convictions in the 1960's. There was no significant change in our reported results (*3SLS* estimates of our model alone also produce no significant difference in results).

II. Estimates of the Structural Model

A. *Estimates of the Model's Coefficients*

Columns (2)–(7) in Table 2 present two-stage least squares (*2SLS*) estimates of the model's coefficients and standard errors. Column (1) shows the single equation estimates of Erhlich's murder supply function using his three-pass least squares technique. Columns (8)–(12) show alternative specifications of the resource equations and the murder supply function. The latter equations are discussed in Section II, Part B.

The coefficients in column (1) indicate that the estimation of Ehrlich's equation using our data and the three-pass technique (compare with equation 3, Table 3, in Ehrlich, 1975a, p. 410) produces results in agreement with Ehrlich's hypotheses regarding the signs and relative sizes of the coefficients of Pa, $Pc|a$, and $Pe|c$. The semilog formulation of the murder supply function estimated using *2SLS* (col. (2)) also supports his hypotheses. The coefficients for Pa, $Pc|a$, and $Pe|c$ are all negative and all more than 1.5 times their standard errors with the coefficient of $Pe|c$ nearly 3 times its standard error.

The estimated production equations for apprehensions, convictions, and executions are largely as expected. The resource variables have positive coefficients. The fraction of the population in the 14–19 age range has a negative coefficient greater than twice its standard error in the equation for executions. The variables related to appeals and lags between conviction and execution also have negative coefficients greater than twice their standard errors in the execution equation. Finally, the coefficients of the workloads Q, A, and C in the production equations have the relative values $0 > a_{22} > a_{32} > a_{42}$ predicted in H-K-S. The importance of this result is discussed in Section III.

The estimates of the resource equations shown in columns (6) and (7) generally support the hypothesis that an increase in the number of crimes leads to a corresponding lagged increase in the amount of resources provided by society to the criminal justice system. The coefficients for the age 25–49 variable are positive and twice their standard errors and provide some support for the hypothesis that persons in this age group are relatively more willing to support crime control activities. However, the positive coefficient for the 50-and-over age group is less than its standard error. This could result from lower incomes in this age bracket. The positive coefficient of the NW variable is consistent with increased criminal justice expenditures over the sample period due to urbanization if, in fact, this variable controls for urbanization.

B. *Tests of Overidentifying Restrictions*

Each of the equations in our structural model is overidentified in terms of the order condition. For each equation we can empirically test the null hypothesis that all of the coefficients of the predetermined variables excluded from the equation are zero against the alternative hypothesis that one or more are not zero. If we let $\lambda = T\hat{\sigma}^2/\hat{\alpha}' W \hat{\alpha}$, then the statistic is $[(T-K)/L](\lambda - 1)$ where $\hat{\sigma}^2$ is the *2SLS* estimate of the residual variance in the equation; $\hat{\alpha}' = (1, -\hat{\alpha}_1')$ where $\hat{\alpha}_1'$ is the vector of *2SLS* estimated coefficients for the endogenous variables on the right-hand side of the equation; W is the matrix of the second moments of the residuals of regressions of the endogenous variables included in the equation on all predetermined variables in the model; T is the number of observations; K is the total number of predetermined variables in the model; and $L = K_2 - G_1$ where K_2 is the number of excluded predetermined variables and G_1 is the number of included endogenous variables on the

TABLE 2—REGRESSION RESULTS FOR THE STRUCTURAL EQUATIONS (1935–69)[a]

	Equations (Number and Dependent Variables)										
	log Q/N (1)	Q/N (2)	log Pa (3)	log Pc|a (4)	log Pe|c (5)	log R (6)	log POL (7)	log R (8)	log POL (9)	Q/N (10)	Q/N (11)
					Jointly Dependent Variables						
log Pa	−2.12	−.10								−.19	−.21
	(.75)	(.044)								(.066)	(.079)
log Pc|a	−.21	−.012								.0045	.010
	(.10)	(.0063)								(.012)	(.015)
log Pe|c	−.056	−.0046								.00016	.00021
	(.022)	(.0016)								(.0033)	(.0039)
log Q			−.13			−.29	−.43	−.62	−.83		
			(.023)			(.12)	(.14)	(.17)	(.20)		
log A				−.55							
				(.26)							
log C					−.93						
					(.57)						
log POL			.15								
			(.022)								
log R				.45	1.52						
				(.34)	(2.54)						
					Predetermined Variables						
log L	−.16	−.032								.046	.068
	(1.05)	(.064)								(.081)	(.11)
log U	.027	.0019								−.0024	−.0030
	(.029)	(.0017)								(.0031)	(.0037)
$\log Y_p$	1.03	.054				−.0055	.052	.17	.27	−.0047	−.022
	(.25)	(.014)				(.28)	(.33)	(.29)	(.35)	(.037)	(.045)
log N						1.04	1.51	−2.09	−2.33		
						(.38)	(.46)	(1.16)	(1.37)		
T	−.041	−.0024								.00041	.00093
	(.0073)	(.00044)								(.0017)	(.0021)
$\log A_1$	.42	.016								.022	
	(.20)	(.013)								(.014)	
$\log A_2$						1.41	2.01	2.96	3.92		
						(.84)	(1.00)	(1.02)	(1.21)		
$\log A_3$						.18	.46	−4.12	−4.83		
						(.81)	(.97)	(1.72)	(2.03)		
$\log A_4$					−2.08						
					(.71)						
$\log A_5$											.016
											(.015)
$\log A_6$										.039	.035
										(.023)	(.024)
log NW			−.61	−.71		3.44	3.76	5.71	6.54		
			(.15)	(1.63)		(1.35)	(1.60)	(1.60)	(1.90)		
CL					−1.75						
					(.32)						
$\log Q_{-1}$						.12	.13	.035	.018		
						(.092)	(.11)	(.10)	(.12)		
$\log Q_{-2}$						.16	.16	.12	.11		
						(.080)	(.095)	(.084)	(.10)		
$T \cdot \log C_{-1}$					−.016			.0085	.010		
					(.0085)			(.0030)	(.0035)		
$\log OC_{-1}$						−.17	−.14	−.18	−.16		
						(.13)	(.15)	(.13)	(.16)		
$\log OC_{-2}$						.20	.25	.17	.22		
						(.10)	(.12)	(.11)	(.13)		
Constant	1.78	.25	3.16	3.50	−6.24	4.53	.26	44.37	49.19	1.08	1.29
	(3.28)	(.20)	(.63)	(7.48)	(21.95)	(4.23)	(5.03)	(14.52)	(17.22)	(.52)	(.64)
R^2	.93	.92	.83	.23	.96	.99	.99	.99	.99	.92	.90
SEE	.047	.0028	.014	.14	.35	.024	.029	.025	.029	.0029	.0033
Value of the Test Statistic[b]	–	4.66	1.99	5.94	1.55	6.07	5.56	1.52	1.22	2.38	1.69
		(8, 18)*	(13, 18)	(13, 18)*	(11, 18)	(6, 18)*	(6, 18)*	(5, 18)	(5, 18)	(8, 17)	(8, 17)

[a] The standard error of the coefficient is in parentheses beneath it.

[b] An asterisk indicates that the value of the test statistic is significant at the 5 percent level.

right-hand side of the equation.[19] This statistic is based on the work of T. W. Anderson and Herman Rubin who provided a test statistic when the estimates are obtained using limited information maximum likelihood techniques. R. L. Basmann derived this statistic for estimates obtained by *2SLS*. Joseph Kadane showed that both the Basmann statistic and the Anderson-Rubin statistic have the F-distribution with L and $T-K$ degrees of freedom.

The statistic can be interpreted as follows. The numerator of λ is the sum of squared residuals of the structural equation restricted by excluding a subset of the predetermined variables. The denominator is the sum of squared residuals with all of the predetermined variables entered in the equation. Hence, a large value of λ indicates that the zero restrictions on the coefficients of the excluded predetermined variables substantially increase the residual variance of the equation. A value of the statistic above the critical value leads to rejection of the null hypothesis in favor of the alternative hypothesis of too many omissions of variables under the null hypothesis. A rejection of the null hypothesis implies that the equation should be reestimated with some of the originally excluded predetermined variables now included. Of course, if this procedure is to improve an equation's specification, any added variables must be justified a priori.

The value of the statistic is given in the last row of Table 2 with the associated degrees of freedom shown in parentheses. From the values of the test statistic we reject the null hypothesis that the overidentifying restrictions on the murder supply function as specified by Ehrlich are correct. We cannot reject the null hypothesis for the production equations for arrests and executions. However, the hypothesis is also rejected for the production equations for convictions and for the resource equations.

For the production equation for convictions we tested the hypothesis that variables entering the equations determining resources should be included for this equation to be specified correctly. When these variables from the resource equation were entered, the value was 2.20 with 4 and 18 degrees of freedom. The hypothesis of correct specification cannot be rejected when the variables from the resource equations are added. However, when we added the predetermined variables from the murder supply function to the conviction equation, the hypothesis of correct specification was rejected.

We tested whether including variables from the murder supply function or the production equations in the resource equations produced test statistic values below the critical value. Only when the interaction term controlling for additional appeals from the execution equation was added was the hypothesis of correct specification not rejected for either resource equation. The inclusion of this variable led us to consider that increased numbers of appeals (which increased for noncapital as well as capital crimes) required additional resources for police who had to develop better cases and for courts and prosecutors who had to deal with heavier workloads.

Columns (8) and (9) of Table 2 show the resource equations estimated with the additional variable. The interaction term variables have positive coefficients greater than twice their standard errors which supports the hypothesis of the additional burden of appeals. The only substantial changes between these equations and those in columns (6) and (7) are the negative coefficients for A_3 and total population. The former could be explained by the negative effect of the reduced income of persons in the latter part of this age group which might offset the positive effect of their expressed fear of crime. The negative coefficient of population is probably due to the high correlation between this variable and others in the equation. The correlations between population and income, the proportion nonwhite in

---

[19] The test statistic we use is correct under the assumption that serial correlation is not present. If serial correlation is present, the statistic is basically of the same form. However, the order condition for each equation is that the number of excluded exogenous (*not* predetermined) variables must be greater than or equal to the number of included current and lagged endogenous variables.

TABLE 3—ARRESTS FOR MURDERS OVER THE SAMPLE PERIOD, 1933–69
(By Age Group)

Age Group	Proportion of Total Arrests for Murder	Age	Proportion
Under 15	0.0077	15	0.0080
15–19	0.1108	16	0.0155
20–24	0.1916	17	0.0231
25–29	0.1737	18	0.0301
30–34	0.1405	19	0.0341
35–39	0.1203	20	0.0346
40–44	0.0886	21	0.0384
45–49	0.0640	22	0.0391
50 and Over	0.1027	23	0.0402
		24	0.0391

Note: The data for this table were taken from the Uniform Crime Reports. The fractions were calculated by summing total arrests for murder from 1933 to 1969 for each age group and dividing each sum by total arrests. The definitions underlying the available data varied as follows: 1933–50—Arrests were for murder, nonnegligent manslaughter, and manslaughter by negligence in all areas of the United States; 1951–63—Arrests were for murder and nonnegligent manslaughter in cities over 2,500; 1964–69—Arrests were for murder and nonnegligent manslaughter in all areas of the United States.

the population, the interaction term, and the lagged values of other violent crimes are all greater than 0.9.

With regard to the murder supply function, the only variables from the production and resource equations that could be justified a priori as belonging in the murder supply function were the age variables A_2, A_3, and A_4 controlling for the age composition of the population.[20] Because these variables cover such a broad segment of the population, we were led to examine murder data by age group according to the *Uniform Crime Reports*.

Table 3 shows the fraction of arrests for murder by five-year age brackets and also for individual ages between 15 and 24. Note that the proportions are high for the age groups 15–19, 20–24, and 25–29, although the more detailed results indicate that the proportion is relatively low for 15-year olds. In fact, for the 15–20-year olds, the detailed results show considerable variation in the proportions of arrests across these years. Another source of variation, not shown in Table 3, is the substantial reduction in arrests of 18- and 19-year olds and of persons in the 20–24 age group during the three wars in the sample period. During these years the proportion of persons arrested in these age groups was well below the average over the sample period.[21]

Given these problems with the data for persons under 25, we decided to enter separate variables in the murder supply function to control for persons under 25 and in the 25–29 age group. (The latter variable is A_6 in Table 2.) Our first regression includes the 14–24 and 25–29 age groups as separate variables, since the data indicate that the proportion of murders committed by persons in the 14–24 age group is high, though variable. Our second regression includes the 21–24 (A_5 in Table 2) and 25–29 age groups separately. Entering only these two variables removes the uneven effect of the 14–20 year olds, but A_5 still includes an uneven effect due to the three wars.

The results of these estimates are shown in columns (10) and (11) of Table 2.[22] Both equations have test statistic values which indicate that the hypothesis of the correctness of the zero restrictions on the coefficients of the excluded exogenous variables cannot be rejected. For the 25–29 age variable the coefficients are about equal in magnitude and greater than their standard errors in both equations. The coefficient of the 14–24 age variable in column (10) is greater than its standard error, while the coefficient of the 21–24 age variable in column (11) is about equal to its standard error. It appears that the effects of the age variables are uneven, probably for the reasons described above.

---

[20] We are indebted to Christopher Sims for suggestions concerning the interpretation of variables controlling for the age composition of the population in the murder supply function.

[21] For example, in 1943 the proportions of arrests were 0.1578 for the 20–24 age group and 0.1749 for the 25–29 year olds.

[22] The estimates of the parameters of the other structural equations were virtually unchanged for the equations in columns (10) and (11). The estimates are available from the authors on request.

In the equations with respecified variables for the age composition of the population, the coefficients of $Pc|a$ and $Pe|c$ are positive and less than their standard errors. In each equation the coefficient of Pa is larger in absolute value than the coefficient of Pa in column (2) and more than twice its standard error. That the coefficient of Pa increases in absolute value seems reasonable to us, for when apprehension occurs, it occurs relatively soon after the crime and thus represents an immediate threat to the murderer. Further, the measured probability of apprehension is large enough so that it is plausible to assume that potential murderers are aware of it as a significant, as well as an immediate, threat. These results suggest that perhaps the true murder supply function is one in which 1) the effect of the probability of apprehension is larger than that found by Ehrlich, and 2) the effects of the conditional probability of conviction given apprehension and execution given conviction are zero.

III. An Interpretation of the Causality Underlying Ehrlich's Specification of the Murder Supply Function

In this section we present an alternative interpretation of Ehrlich's empirical results. He found that the coefficients for the deterrence variables were negative, larger than their standard errors, and in the ascending order predicted by his theory. We suggest that these impressive results could have been caused by the response of the criminal justice system to murders. Two questions must be answered to determine the plausibility of this alternative interpretation. First, does the theory of the response of society and the criminal justice system to murder behavior provide the same prediction of ascending coefficients as does Ehrlich's theory of murder behavior? Second, do the empirical results for the response of society and the criminal justice system support the alternative interpretation?

The simultaneous solution of the production equations and resource equations (2)–(6) is shown by H-K-S to yield the equation:

$$(10)\quad \log Q = K_0 + K_1 \log Pa + K_2 \log Pc|a + K_3 \log Pe|c + \sum_{i=4}^{n} K_i \log V_{i-3} + \varepsilon$$

where the V_is are the predetermined variables in the production and resource equations and the K_is are the associated coefficients, and

$$(11)\quad K_1 = 1 \div 3(a_{23}a_{62} + a_{22})$$

$$(12)\quad K_2 = 1 \div 3\left[(a_{33}a_{52} + a_{32}) + a_{32}(a_{23}a_{62} + a_{22})\right]$$

$$(13)\quad K_3 = 1 \div 3\left[(a_{43}a_{52} + a_{42}) + a_{42}(a_{23}a_{62} + a_{22} + (a_{33}a_{52} + a_{32}) + a_{32}(a_{23}a_{62} + a_{22}))\right]$$

With the exogenous variables used by Ehrlich in his murder supply function entered into the production or resource equations, equation (10) is identical to Ehrlich's murder supply function. However, the causality underlying this equation is the response of the punishments meted out by the criminal justice system to murder behavior. While this equation's particular way of expressing this response is not of intrinsic interest, it does illustrate that Ehrlich's murder supply function, if not identified, could represent criminal justice system behavior and not murder behavior. In particular, K_1, K_2, and K_3 represent the effects of murders on Pa, $Pc|a$, and $Pe|c$. A given increase in murders adds to the criminal justice system's workloads but does not add to its resources to deal with these workloads except with lags. Thus, K_1, K_2, and K_3 can be expected to be negative. The coefficient K_2 has more terms than K_1 because the response of the apprehension rate to increased murders is part of the response of the conviction rate to increased murders, since apprehensions are the workload for the part of the criminal justice system dealing with convictions. Similarly, K_3 has more terms than K_2 because the response of the conviction rate to increased murders is part of the response of the execution rate to increased murders. Of particular interest are the relative magnitudes of K_1, K_2 and K_3. Hoenack, Kudrle, and Sjoquist derive sufficient conditions for $K_1 < K_2 < K_3 < 0$; the condition having the most behavioral significance is that $a_{42} < a_{32} < a_{22} < 0$.[23] Section I points out that the H-K-S theory provides the a priori prediction $a_{42} < a_{32} < a_{22} < 0$. The estimates for the production equations in Section II show this result.

We now present estimates of the values of the solved coefficients in equation (10) under alternative specifications of the estimated production and resource equations from which it is derived.

We estimated equations (2)–(6) using ordinary least squares and calculated the values of K_1, K_2, and K_3 under two sets of assumptions.[24] Under the first set of assumptions, we respecified equations (2)–(6) in such a way that only the exogenous variables in Ehrlich's equation were used in the solution for equation (10). In particular, we specified the production equations for $Pa, Pc|a$, and the execution rate (we used both the $PXQ1_{-1}$ and $PXQ2$ execution risk measures defined by Ehrlich) with only Q and POL on the right-hand side of (2), A and R on the right-hand side of (3), and C and R on the right-hand side of (4). The exogenous variables L, U, T, A_1, and Y_p were included on the right-hand side of (5) and (6). In the first variant under these assumptions we specified the equations (2) for Pa and (5) and (6) for resources with Q on the right-hand side, and in the second variant Q/N was substituted for Q in these equations. If Q/N is on the right-hand side of (2), (5), and (6), the solved equation (10) is identical to that estimated by Ehrlich. The first two rows in Table 4 present the corresponding values of K_1, K_2, and K_3 calculated from equations (11)–(13). While these coefficients are somewhat different in magnitude from those that Ehrlich found, each row has $K_1 < K_2 < K_3 < 0$, which is the condition he predicted and found.

Under the second set of assumptions, we specified the equations for Pa, $Pc|a$, the execution rate, and resources exactly as shown in equations (2)–(6). The first variant under this set of assumptions included Q on the right-hand side of equations (2), (5), and (6); the second variant substituted Q/N for Q in these equations. The third and fourth rows in Table 4 show the solved values of K_1, K_2, and K_3 for this set of equations. While different in magnitude from the coefficients in the first two rows of Table 4, the

TABLE 4—SOLVED COEFFICIENTS OF EQUATION (10)

Dependent Variables	K_1	K_2	K_3	
			$PXQ1_{-1}$	$PXQ2$
Q	-1.78	$-.68$	$-.39$	$-.37$
Q/N	-1.69	$-.68$	$-.41$	$-.39$
Q	-1.66	$-.59$	$-.42$	$-.51$
Q/N	-1.88	$-.57$	$-.42$	$-.50$

Note: The regressions used to calculate these values are available from the authors on request.

---

[23] The conditions are

(a) equations (2)–(4) possess constant returns to scale, i.e.,

$$a_{22}+a_{23}+1=a_{32}+a_{33}+1=a_{42}+a_{43}+1=1$$

(b) $a_{42}, a_{62} < 1$

(c) $a_{23} < \dfrac{a_{33}}{1+a_{33}}$

(d) $a_{33} < \dfrac{a_{43}}{1+a_{43}}$

(e) $0 > a_{22} > a_{32} > a_{42}$

[24] For the purposes of these calculations we treated Q as predetermined or determined only by exogenous variables. In the results presented here the former condition is assumed to hold. If the latter condition were to hold, the results would be modified only slightly.

coefficients in the third and fourth rows again maintain the ordering $K_1 < K_2 < K_3 < 0$.

The results in Table 4 suggest that Ehrlich's estimated equation could have been causally generated by the response of the criminal justice system to murder. In this alternative interpretation, Ehrlich's estimated equation represents production behavior within the criminal justice system and the provision of resources to it, although misspecified. Is this alternative interpretation more plausible than the murder supply interpretation? The alternative interpretation is supported by the failure of the estimated murder supply function with ascending negative coefficients larger than their standard errors for the three deterrence probabilities to pass the test for correctness of overidentifying restrictions. On the other hand, when the respecified murder supply function passes the test for correctness of overidentifying restrictions, the ascending negative coefficient result vanishes.[25]

IV. Conclusion

Because of the high correlation among most of the time-series data used in this study, we do not regard our results as at all conclusive. The potential policy significance of deterrence estimates could justify the necessary expenditures in collecting data for cross-section estimation of a structural model. These data should encompass actual capital murders and the specific criminal justice activities associated with them. Equally important, however, is the need for improved theory. Ehrlich's murder supply function contained eight variables, three of which were justified by his deterrence theory and five of which were given much less justification. We have seen that an improved understanding of the response of the criminal justice system to crime can lead to an improved specification of criminal behavior as well as to an improved specification of equations for society's response to crime. Our initial attempt at a structural model, for example, led through an examination of the correctness of overidentifying restrictions to an improved explanation of the inclusion of variables controlling for the age composition of the population in the murder supply function and of a variable controlling for increased appeal activity in the resource equations. We have provided an example that illustrates the necessity of having a theory about all of the equations in a structural model and about every variable in each of its equations.

APPENDIX

Researchers critical of Ehrlich's work have attempted to replicate his data closely. We had planned to replicate his data using the modifications suggested by W. J. Bowers and G. L. Pierce.[26] However, on close investigation of the data, we decided that it would be preferable to construct a data base with internally consistent definitions and with the revised data now available. Our reasons were as follows. First, the data for murders and other violent crimes reported through the Uniform Crime Reporting Program by law enforcement agencies only include murders investigated by the police within the United States. Thus the other variables should include only data for civilians within the United States. Ehrlich's civilian population variable (as well as his other variables which include population) includes Alaska and Hawaii before they became states. His population and labor force variables also include members of the armed forces overseas in several years.

---

[25] Also supporting the alternative interpretation is Ehrlich's showing that the *log*-linear form of his equation is superior in his application of the Box and Cox procedure for selection of optimal data transformations. (See Erhlich, 1977a, p. 454.) While the *log*-linear form is straightforwardly derived from a Cobb-Douglas specification of the production equations for the criminal justice system, it would require an awkward application of Ehrlich's theory of rational murder behavior to both capital and total murders. It is also noteworthy that in Ehrlich's interpretation murderers are required to respond to small and distant probabilities. In the alternative interpretation, the coefficient for each of these probabilities is determined by the constrained response of the criminal justice system to changes in workloads.

[26] We are grateful to Glenn Pierce for making available to us the Bowers and Pierce data and Ehrlich's memorandum describing his data sources (see Ehrlich and Mark, 1975).

Second, while Ehrlich used the latest data on number of murders available at the time he wrote his article, these numbers have been revised since 1970.[27] These revised data also change the estimated numbers of arrests, convictions, and via convictions, the estimated execution rate. Third, the population data available to Ehrlich for the 1960's were estimates based on the 1960 Census. These data have been revised based on the 1970 Census, and we used the revised data. Fourth, our population variables (civilian population, proportion nonwhite, proportions in various age groups) are based on civilian populations, excluding armed forces overseas and/or stationed in the United States, and excluding Alaska and Hawaii before 1959. Fifth, our labor force participation rate is the ratio of the civilian labor force aged 14 and above to the civilian population of the same age. Ehrlich used the rate for those aged 16 and above for 1947 and beyond, and the rate for those aged 14 and over before 1947. A detailed listing of our data sources and the actual data are available on request.

---

[27] The reported number of murders is higher in each year in the sample period (1933–69). The mean number of murders per year over the sample period in the data used by Ehrlich is 8,946; in our data the mean is 9,006.

REFERENCES

T. W. Anderson and H. Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals Math. Statist.*, Mar. 1949, *20*, 46–63.

D. C. Baldus and J. W. L. Cole, "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment," *Yale Law J.*, Dec. 1975, *85*, 170–86.

R. L. Basmann, "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *J. Amer. Statist. Assn.*, Dec. 1960, *55*, 650–59.

Hugo A. Bedau, *The Courts, the Constitution, and Capital Punishment*, Lexington 1977.

Monica D. Blumenthal et al., *Justifying Violence: Attitudes of American Men*, Ann Arbor 1972.

Donald J. Bogue, *The Population of the United States*, New York 1959.

W. J. Bowers and G. L. Pierce, "The Illusion of Deterrence in Isaac Ehrlich's Research on Capital Punishment," *Yale Law J.*, Dec. 1975, *85*, 187–208.

O. G. Brim, Jr., "Adult Socialization" in John A. Clausen, ed., *Socialization and Society*, Boston 1968.

Hugh Carter and Paul C. Glick, *Marriage and Divorce: A Social and Economic Study*, Cambridge, Mass. 1970.

J. L. Chapman, "The Demand for Police," *Publ. Finance Quart.*, Apr. 1976, *4*, 187–204.

Caryl Chessman, *Cell 2455, Death Row*, New York 1954.

______, *Trial By Ordeal*, New York 1955.

______, *The Face of Justice*, Englewood Cliffs 1957.

C. T. Clotfelter, "Public Spending for Higher Education: An Empirical Test of Two Hypotheses," *Publ. Finance*, No. 2, 1976, *31*, 177–95.

P. J. Cook, "Punishment and Crime: A Critique of Current Findings Concerning the Preventive Effects of Punishment," work. paper OPP-ERP 76-3, 22–40, U.S. Dept. Justice, Aug. 1976.

I. Ehrlich, "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *J. Polit. Econ.*, May/June 1973, *81*, 521–65.

______, (1975a) "The Deterrent Effect of Capital Punishment: A Question of Life and Death," *Amer. Econ. Rev.*, June 1975, *65*, 397–417.

______, (1975b) "Deterrence: Evidence and Inference," *Yale Law J.*, Dec. 1975, *85*, 209–27.

______, (1977a) "The Deterrent Effect of Capital Punishment: Reply," *Amer. Econ. Rev.*, June 1977, *67*, 452–58.

______, (1977b) "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence," *J. Polit. Econ.*, Aug. 1977, *85*, 741–88.

______ **and J. C. Gibbons,** "On the Measurement of the Deterrent Effect of Capital Punishment and the Theory of Deterrence," *J. Legal Stud.*, Jan. 1977, *6*, 35–50.

______ **and R. Mark,** "Sources of Data for 'The Deterrent Effect of Capital Punishment: A Question of Life and Death'," unpublished memo., rev. Sept. 1975.

______ **and** ______, "Fear of Deterrence: A Critical Evaluation of the 'Report of the Panel on Research on Deterrent and Incapacitative Effects'," *J. Legal Stud.*, June 1977, *6*, 293–316.

H. Erskine, "The Polls: Fear of Violence and Crime," *Publ. Opinion Quart.*, Spring 1974, *38*, 131–45.

F. M. Fisher and D. Nagin, "On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime and Sanctions," in Alfred Blumstein, ed., *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, Washington 1977.

B. Forst, V. Filatov, and L. R. Klein, "The Deterrent Effect of Capital Punishment: An Assessment of the Estimates," in Alfred Blumstein, ed., *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, Washington 1977.

L. S. Friedman, "The Use of Multiple Regression Analysis to Test for a Deterrent Effect of Capital Punishment: Prospects and Problems," mimeo., Grad. Sch. Publ. Pol., Univ. California-Berkeley 1975.

Paul C. Glick, *American Families*, New York 1957.

S. A. Hoenack, R. T. Kudrle, and D. L. Sjoquist, "The Deterrent Effect of Capital Punishment: A Question of Identification," *Pol. Anal.*, Fall 1978, *4*, 491–527.

J. B. Kadane, "Testing Overidentifying Restrictions When the Disturbances Are Small," *J. Amer. Statist. Assn.*, Mar. 1970, *75*, 182–85.

W. M. Landes, "An Economic Analysis of the Courts," *J. Law Econ.*, Apr. 1971, *14*, 61–107.

E. J. Mishan, "Evaluation of Life and Limb: A Theoretical Approach," *J. Polit. Econ.*, July/Aug. 1971, *79*, 687–705.

J. T. Mortimer and R. G. Simmons, "Adult Socialization," in Ralph H. Turner et al., eds., *Annual Review of Sociology*, Vol. 4, Palo Alto 1978.

P. Passell and J. B. Taylor, "The Deterrent Effect of Capital Punishment: Another View," *Amer. Econ. Rev.*, June 1977, *67*, 445–51.

L. Phillips and H. L. Votey, "Crime Control in California," *J. Legal Stud.*, June 1975, *4*, 327–49.

R. Thaler, "An Econometric Analysis of Property Crime," *J. Publ. Econ.*, Aug. 1977, *8*, 37–51.

Burton H. Wolfe, *Pileup on Death Row*, Garden City 1973.

Adderly v. Wainwright, 385 U.S. 39 (1967).

Brown et al. v. State of Mississippi, 297 U.S. 278 (1936).

Escobedo v. State of Illinois, 378 U.S. 478 (1964).

Hill, Saterfield and Anderson v. Nelson, 272 F. Supp 790 (1967).

Mallory v. United States, 354 U.S. 449 (1957).

Norris v. Alabama, 294 U.S. 587 (1935).

People v. Dorado, 42 Cal Rptr 169 (1965).

People v. Morse, 388 P. 2d 33 (1964).

U.S. Bureau of the Census, *Statistical Abstract of the United States*, Washington 1974.

U.S. Department of Justice, Federal Bureau of Investigation, *Crime in the United States: Uniform Crime Report*, Washington, various issues.

Capital Contracting and the Regulated Firm

By H. STUART BURNESS, W. DAVID MONTGOMERY, AND JAMES P. QUIRK*

Choices by firms between risky and safe options, as in contracting for capital goods, are usually discussed by economists in terms of the attitudes towards risk on the part of supplying and/or demanding firms. Other aspects of this problem arise in the case of regulated firms. The structure of public utility regulation raises questions as to whether incentives are created for utilities to bear capital costs risks. A case in point occurred during the so-called "turnkey era" in nuclear power. During the turnkey era from 1963 to 1966 the leading *U.S.* reactor manufacturers (General Electric and Westinghouse) offered utilities turnkey (fixed-price) contracts covering essentially all the costs of nuclear reactor construction, presumably in an effort to generate interest (and sales) in the fledgling nuclear power industry.[1] Even during the turnkey era, roughly one-third of the reactors ordered were on nonturnkey terms and immediately following the turnkey era, the return to conventional cost-plus contracting was followed by a surge in orders for nuclear reactors. A partial explanation might be that utilities prefer bearing the cost risks of capital contracting to fixed-price contracts.

Consequently, we examine risk-sharing arrangements involving regulated firms, first in the traditional Averch-Johnson (hereafter A-J) framework and subsequently in the context of a variant of a model of regulation developed by Paul Joskow. Under the simplifying assumption of expected income maximization, it is shown that utilities prefer a turnkey contract to a conventional contract in the A-J framework; but this conclusion is reversed using a Joskow-type model. As Joskow has emphasized, there are basic problems with the A-J model, particularly during periods such as the 1960's and 1970's, when the impetus for regulatory review comes from the utility rather than from the regulating agency. Our final section summarizes utility-purchasing behavior and regulatory activity during and immediately following the turnkey era and compares this with the theoretical results derived in the body of the paper.

I. The A-J Model and Risk Sharing

We consider a variant of the A-J model in which the allowed profits of the regulated utility are based on the observed cost to the utility per unit of capital. The utility has choices with respect to the number of units of capital K, and the number of units of some index of variable inputs L to purchase.[2] The price of variable inputs w is known with certainty, but the cost per unit of capital to the supplier of capital p is a random variable with known probability distribution $f(p)$.

Under a turnkey contract, the utility can buy as many units of capital as it wishes at a guaranteed price $E(p) \equiv \int_0^\infty pf(p)dp$ per unit of capital. Let Π denote profits and let $R(K,L)$ denote revenue to the utility, with R strictly concave and monotonic increasing in K,L, where $\lim_{K\to\infty} R_K = \lim_{L\to\infty} R_L = 0$. Then the utility solves the problem

$$\max_{K,L} \Pi = R(K,L) - wL - rE(p)K$$

subject to $$R(K,L) - wL \leqq sE(p)K$$

---

*University of New Mexico, U.S. Department of Energy, and California Institute of Technology, respectively. This research was supported in part under NSF APR75-16566 and under ERDA EY-76-G-03-1305. We wish to thank the Environmental Quality Laboratory at Caltech for its assistance, and Roger G. Noll, Ron Braeutigam, and Paul Joskow for instructive comments, as well as Cindy Carlson for her help with the data used in this paper. We are solely responsible for all opinions and conclusions.

[1] For a detailed expositon and analysis of the turnkey era and utility-reactor manufacturer behavior during this time see our paper.

[2] The most important variable input is fuel; operation and maintenance labor can also be varied, and a utility can purchase power for resale if its capital stock K imposes a constraint on generating capacity.

where r is the market rate of return per unit of capital and s is the allowed rate of return, $r<s<r_M$, with r_M being the rate of return for an unregulated monopolist.

Let $G=\Pi+\lambda(sE(p)K-R(K,L)+wL)$. At a constrained maximum of profits under a turnkey contract we have

$$(1)\quad R_K(K_T,L(K_T))=\frac{E(p)(r-\lambda s)}{(1-\lambda)}$$

$$(2)\quad R_L(K_T,L(K_T))=w$$

$$(3)\quad sE(p)K_T=R(K_T,L(K_T))-wL(K_T)$$

where K_T is the optimizing value of K, and $L(K_T)$ is the value of L that satisfies (2).

In (1), $r<s<r_M$ with R strictly concave implies $0<\lambda<1$; in fact, it is clear from (1) that $0<\lambda<r/s<1$. Further, since (3) holds as an equality, we have

$$\Pi_T(K_T)=(s-r)E(p)K_T$$

where $\Pi_T(K_T)$ denotes the maximum value of profits under a turnkey contract, given that the utility is subject to rate-of-return regulation.

When we turn to the case of a conventional contract for the purchase of capital goods, things are somewhat more complicated. The scenario we assume is the following. The utility must choose the value of K before p is observed. Once K is chosen and p is observed, then the utility chooses L in a profit-maximizing fashion, subject to the rate-of-return constraint.

As above, for any K, define $L(K)$ as that value of L that satisfies $R_L(K,L(K))=w$. $L(K)$ is the value of L that would be chosen, given K, if there were no rate-of-return constraint restricting choices.

Given K, let $p^*(K)$ denote the value of p that satisfies

$$p^*(K)=\frac{R(K,L(K))-wL(K)}{sK}$$

For $p\geqslant p^*(K)$, the rate-of-return constraint is ineffective and $L=L(K)$, while if $p\leqslant p^*(K)$, the constraint is effective and $L=L^*(K)$, where $L^*(K)$ satisfies $R[K,L^*(K)]-wL^*(K)-rpK=spK$. Profits under a conventional contract then are given by

$$\Pi_C(p,K)=\begin{cases} R(K,L(K))-wL(K)-rpK \\ \text{for}\quad R(K,L(K)) \\ \qquad -wL(K)\leqq spK \\ (s-r)pK \\ \text{for } R(K,L(K))-wL(K)\geqq spK \end{cases}$$

In brief, for any given K, profits are the maximum possible for that K (i.e., $L=L(K)$) so long as p is large enough so that the rate-of-return constraint is not binding, i.e., $p\geqq p^*(K)$. If p is so small that choosing $L=L(K)$ would violate the rate of return constraint, i.e., $p\leqq p^*(K)$, then L is chosen so that profits equal the maximum permitted under rate-of-return regulation.

Rewriting, we have

$$\Pi_C(p,K)=\begin{cases} R(K,L(K))-wL(K)-rpK \\ \qquad\qquad \text{for } p\geqq p^*(K) \\ (s-r)pK \text{ for } p\leqq p^*(K) \end{cases}$$

An expected income-maximizing utility then chooses K to solve the problem.

$$\max_K \int_0^{p^*(K)}(s-r)pKf(p)dp+$$
$$\int_{p^*(K)}^{\infty}\left[R(K,L(K))-wL(K)-rpK\right]f(p)dp$$

with first-order condition requiring that at the optimizing K, K_C,

$$(4)\quad \int_0^{p^*(K_C)}(s-r)pf(p)dp + \int_{p^*(K_C)}^{\infty}\left[R_K-rp\right]f(p)dp=0$$

Rewriting, we have

$$(4')\quad R_K(K_C,L(K_C))=\frac{rE(p)-s\int_0^{p^*(K_C)}pf(p)dp}{(1-F(p^*))}$$

where K_C is the optimizing choice of K.

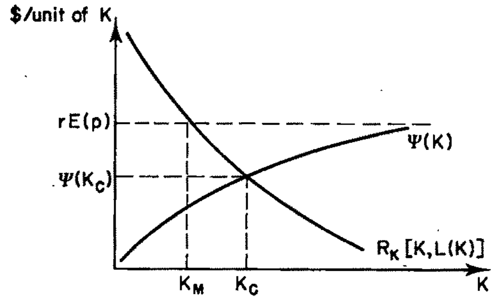


FIGURE 1. $K_C > K_M$

We first show that $K_M < K_C$ for $r < s < r_M$, where K_M is the value of K that would be chosen by an unregulated monopolist. Strict concavity of R in K, L implies $dR_K/dK < 0$.[3] In (4′) let

$$\psi(K) = \frac{rE(p) - s\int_0^{p^*} pf(p)\,dp}{(1 - F(p^*))}$$

Then $$\Psi'(K) = \left(\frac{dp^*}{dK}\right)\left(\frac{f(p^*)}{(1-F(p^*))^2}\right) \times \left(\frac{-E\Pi_C(K)}{K}\right)$$

with $$\frac{dp^*}{dK} = \frac{-E\Pi_C(K)}{[1-F(p^*)]sK^2} \leqslant 0$$

Hence $\Psi'(K) \geqslant 0$, with $\Psi'(K_M) > 0$. Moreover, $\lim_{K\to\infty} \Psi(K) = rE(p)$, while $R_K(K_M, L(K_M)) = rE(p)$ where K_M is the choice of capital by the unregulated monopolist. Hence the graphs of $R_K, \Psi(K)$ appear as in Figure 1 with $K_C > K_M$.

We next show that $K_C < K_T$. Under a conventional contract, profits Π_C are given by

$$\Pi_C(p,K) = \begin{cases} R(K,L(K)) - wL(K) - rpK & \text{for } p \geqq p^*(K) \\ (s-r)pK & \text{for } p \leqq p^*(K) \end{cases}$$

Thus

$$\frac{\partial \Pi_C(p,K)}{\partial K} \equiv h(p,K) = \begin{cases} R_K(K,L(K)) - rp & \text{for } p \geqq p^*(K) \\ (s-r)p & \text{for } p \leqq p^*(K) \end{cases}$$

so that

$$\frac{\partial h(p,K)}{\partial p} \equiv \frac{\partial^2 \Pi_C}{\partial K \partial p} = \begin{cases} -r & \text{for } p > p^*(K) \\ (s-r) & \text{for } p < p^*(K) \end{cases}$$

with $\partial^2 h(p,K)/\partial p^2 = 0$, $p \neq p^*(K)$. Hence $h(p,K)$ is concave in p as shown in Figure 2.

Consequently, $E\{h(p,K)\} < h\{E(p),K\}$ for any $K > 0$, assuming $f(p)$ is nondegenerate with $f(p_1) > 0$; $f(p_2) > 0$ for some $p_1 < p^*(K) < p_2$; otherwise $E\{h(p,K)\} = h\{E(p),K\}$.

The optimizing rule under conventional contracting is to choose $K = K_C$ such that $E\{h(p,K_C)\} = 0$; that is, the expected value of marginal profitability of capital equals zero. Under a turnkey contract, the optimizing rule is to choose $K = K_T$ such that $h(E(p),K_T) = 0$; that is, the marginal profitability of capital at its expected price equals zero. To see this, recall that

$$h(p,K) = \begin{cases} R_K(K,L(K)) - rp & \text{for } p \geqq p^*(K) \\ (s-r)p & \text{for } p \leqq p^*(K) \end{cases}$$

Thus

$$h(E(p),K) = \begin{cases} R_K(K,L(K)) - rE(p) & \text{for } E(p) \geqq p^*(K) \\ (s-r)E(p) & \text{for } E(p) \leqq p^*(K) \end{cases}$$

Since, under a turnkey contract, the utility solves the problem $\max \Pi_T = R(K,L) - wL - rE(p)K$, subject to $sE(p)K - R(K,L) - wL \geqq 0$ with solution $\Pi_T = R(K_T, L(K_T))$

---

[3] This is easy to see as

$$\frac{dR_K(K,L(K))}{dK} = R_{KK} + R_{KL}\frac{dL(K)}{dK}$$

where $\frac{dL(K)}{dK} = \frac{-R_{KL}}{R_{LL}}$ from $R_L(K,L(K)) = w$

Hence $$\frac{dR_K(K,L(K))}{dK} = \frac{R_{KK}R_{LL} - R_{KL}^2}{R_{LL}} < 0$$

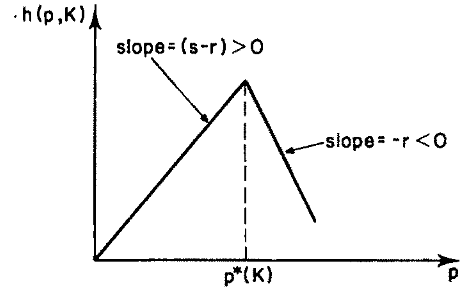


FIGURE 2. $h(p,K)$ FOR K FIXED

$-wL(K_T)-rE(p)K_T=(s-r)E(p)K_T$, it follows that for $K \leqq K_T, (E(p) \leqq p^*(K))$, the constraint is binding and marginal profitability is $(s-r)E(p)$; for $K \geqq K_T(E(p) \geqq p^*(K))$, the constraint is not binding and marginal profitability is $R_K(K, L(K)) - rE(p)$.

Hence under a turnkey contract, K is chosen so that $K=K_T$ with K_T satisfying $h(E(p), K_T)=0$. Since $dh(p,K)/dK<0$, $E\{h(p,K)\}<h(E(p),K)$ for any K implies $K_C<K_T$ as shown in Figure 3; that is, less capital is ordered under conventional contracting than under a turnkey contract.

Recalling the definition of $p^*(K)$ and $E\Pi_C(K)$ we have $E\Pi_C(K)=K\{s[p^*(1-F(p^*))+\int_0^{p^*} pf(p)dp]\}-rE(p)$. Since $p^*[1-F(p^*)]+\int_0^{p^*} pf(p)dp<E(p)$ for $F(p^*)<1$, it follows that $E\Pi_C(K) < \Pi_T(K) \equiv (s-r)E(p)K$. Since $K_C<K_T$, we have finally $E\Pi_C(K_C)<\Pi_T(K_T)$ so that the utility prefers the turnkey contract.[4]

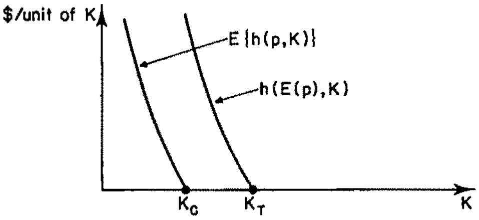


FIGURE 3. $E\{h(p,K)\}$ AND $h(E(p),K)$

PROPOSITION 1: *Given an A-J model of the regulated firm, with revenue strictly concave in K,L, and with $r<s<r_M$, an expected income-maximizing utility will order more units of capital under a turnkey contract than under a conventional contract, and profits under a turnkey contract will be higher than expected profits under a conventional contract. If $s=r$ or $s=r_M$, the utility is indifferent between a turnkey contract and a conventional contract.*

Turnkey contracts are almost unprecedented in the *U.S.* electric utility industry; nonetheless the A-J model implies that expected income-maximizing utilities would be willing to pay a premium for such contracts. To square this result with the history of the electric utility industry would require the assumption that contractors are risk averse, and in fact are more risk averse (in some sense) than utilities, an assumption that lacks much in the way of intuitive appeal.

What this exercise suggests is that there are problems with at least the version of the A-J model that we have employed, and in fact it is not difficult to identify what those problems are. In the A-J model, there is only rate-of-return regulation and it is assumed to be always binding (in the certainty case). But in fact there are two inequalities of relevance in regulated industries, namely,

$$R(K,L)-wL \leqq spK$$

and

$$R(K,L)-wL \geqq rpK \quad ,$$

When utilities are observed to violate the first constraint then it is the Public Utilities

[4] This assumes $r<s<r_M$. If $s=r_M$, then the rate-of-return constraint is no longer binding so that $p^*=0$ and $\lambda=0$. The problem under a conventional contract reduces to

$$\max_K \int_0^\infty (R(K,L(K))-wL(K)-rpK)f(p)dp,$$

with $$R_K(K_C,L(K_C))=rE(p)$$

Thus when $s=r_M$, the utility sets $K_C=K_T$, and expected profits are the same under a conventional contract as under a turnkey contract. If $s=r$, the optimal choice under a conventional contract is to choose $K_C=0$ since profits can never be positive and will be negative for some p if $F(p^*)<1$ with $K>0$. Hence expected profits are zero with $K_C=0$. On the other hand, $\Pi_T(K_T)=0$ for every K_T when $s=r$. Hence the utility is indifferent between a conventional contract and a turnkey contract when $s=r$.

Commission (PUC) or intervenor groups that institute regulatory review; but when the second constraint (earning at least the market rate of return on capital) is violated, it is the utility that institutes the review process. Increasingly over the past ten to fifteen years, it has been the second constraint rather than the first that has been relevant to utility decision making as Joskow has documented.

Moreover, rate-of-return regulation is almost always accompanied by output-pricing regulation, so that the utility in fact is not able to choose its revenue freely subject only to rate-of-return considerations. The existence of output-price regulations coupled with common carrier requirements provides an explanation for the observed fact that electric utilities operate in the inelastic portion of their demand curves. We attempt to incorporate some of these considerations into an admittedly highly simplified model of a "Joskow-type" regulated utility in the next section.

II. A Joskow-Type Model of the Regulated Utility

The variant of the Joskow model that we employ incorporates the following basic assumptions:

1) the price of output P is set by the regulatory agency;

2) output X is given by $X=F(K,L)$, where F is characterized by increasing returns to scale;

3) the utility must meet the common carrier requirement, $X \geqq D(P)$, where $D(P)$ is demand at the output price P.

Profits for the utility are given by

$$\Pi = PD(P) - wL - rpK$$

In general, the price-setting rules for the regulatory commission will depend upon both the rate-of-return constraint and the nonnegativity of profits constraint. We consider only a simplified version of this model, namely one in which regulatory review is initiated by the utility only, so that the rate-of-return constraint does not operate to limit profits. Under this simplified version of the Joskow model, P is set according to the rules

$$P = \begin{cases} P_0 & \text{if } \Pi(P_0, p, K, L) \geqq 0 \\ P_1(p,K,L) & \text{if } \Pi(P_0, p, K, L) \leqq 0 \end{cases}$$

where P_1 is defined by $\Pi(P_1(p,K,L),K,L)=0$.[5]

Thus it is assumed that there is some historically determined P_0, perhaps based on rate-of-return considerations using values of p and K from some test period, which will remain in force so long as the utility earns at least the market rate of return r on its capital. If however, the utility, given P_0, would show a rate of return on its capital less than r, then the utility institutes a rate review and obtains a rate hike just sufficient to enable it to earn the rate of return r.

We begin our analysis of this simple model with the assumption that, for the relevant range of all variables, $D(P)$ is inelastic. Thus the utility is always able to increase revenue by increasing P. Clearly this is unrealistic for sufficiently large values of P, and the conclusions derived for the simple model have to be reexamined when the possibility of an elastic segment of the demand curve occurring is admitted. It is also unrealistic to assume that the PUC would stand still for extremely high profit levels; the rate-of-return constraint certainly comes into play for profits sufficiently large. But we ignore this complication in the present paper.

Under a turnkey contract, the utility solves the problem $\max_{K,L}\Pi = P_0D(P_0) - wL - rE(p)K$ subject to the constraint $F(K,L) \geqq D(P_0)$. Let H denote the Lagrangian, where $H = \Pi + \lambda(F(K,L) - D(P_0))$. Then the first-order conditions are given by

$$\lambda F_K(K_T, L_T) - rE(p) = 0$$

$$\lambda F_L(K_T, L_T) - w = 0$$

$$F(K_T, L_T) = D(P_0)$$

---

[5] We purposely understate our case here as one would expect utility-initiated proceedings to be directed towards a rate hike that would generate the maximum allowable rate of return other than merely zero profits. However, recent regulatory agency decisions of a less-lenient nature suggest a more cautious analytical approach.

where K_T, L_T are the optimizing choices of K, L. Define $L(P,K)$ by $F[K, L(P,K)] = D(P)$ and let $\Pi_T(K_T)$ denote the maximum value of Π subject to the common carrier constraint. Then[6]

$$\Pi_T(K_T) = P_0 D(P_0) - wL(P_0, K_T) - rE(p)K_T$$

If the utility uses conventional contracting procedures, then its decision problem is more complex because it must choose K before the price per unit of capital, p, is known. Then it chooses L to meet demand. Thus given P_0 and K, $L(P_0, K)$ is the number of units of variable inputs required to meet the common carrier requirement for the utility.

For any given K, define $P^0(K)$ as that price per unit of capital that satisfies

$$P_0 D(P_0) - wL(P_0, K) - rp^0(K)K = 0$$

That is, for any choice of K and for the implied choice of L to meet demand $D(P_0)$, $p^0(K)$ is the price of capital at which P_0 equals average cost so that profits above the competitive return on capital, rpK, are zero. If the observed price of capital exceeds $p^0(K)$, then the utility initiates review proceedings and has its rate adjusted to P_1, the output price which equals average cost for the observed p. Thus an expected income-maximizing utility solves the problem

$$\max_K E\Pi_C = \int_0^{p^0} \{P_0^{(K)} D(P_0) - wL(P_0, K) - rpK\} f(p)\,dp + \int_{p^0(K)}^{\infty} \{P_1(p,K,L) D(P_1) - wL(P_1, K) - rpK\} f(p)\,dp$$

Since the second integrand is identically zero for every $p \geqslant p^0(K)$, the first-order condition reduces to

$$\int_0^{p^0(K_C)} \left[-w \frac{\partial L(P_0, K_C)}{\partial K} - rp \right] f(p)\,dp = 0$$

where K_C is the optimizing choice of K under conventional contracting. Thus, the maximum value of $E\Pi_C$, denoted by $E\Pi_C(K_C)$, is given by

$$E\Pi_C(K_C) = \int_0^{p^0(K_C)} \{P_0 D(P_0) - wL(P_0, K_C) - rpK_C\} f(p)\,dp$$

A direct argument establishes that $E\Pi_C(K_C) > \Pi_T(K_T)$ so that the utility prefers conventional contracting to a turnkey contract. Assume that under a conventional contract the utility employed the same stock of capital as was optimal under a turnkey contract. Then we have

$$E\Pi_C(K_T) - \Pi_T(K_T) = \int_{p^0(K_T)}^{\infty} [P_1 D(P_1) - P_0 D(P_0)] f(p)\,dp + w\int_{p^0(K_T)}^{\infty} [L(P_0, K_T) - L(P_1, K_T)] f(p)\,dp$$

Note $p > p^0(K_T)$ implies $P_1 > P_0$; with inelastic demand, $P_1 D(P_1) > P_0 D(P_0)$ for all $p > p^0(K_T)$. Similarly, since

$$D(P_1) < D(P_0) \quad \text{for} \quad p > p^0(K_T),$$
$$L(P_0, K_T) > L(P_1, K_T)$$

Therefore, $E\Pi_C(K_T) > \Pi_T(K_T)$; that is, expected profits are higher under conventional contracting than under turnkey contracting when the utility chooses the same K in both cases. Because the utility will choose $K_C \neq K_T$ only if the change increases expected profits, it follows that conventional contracting always provides expected profits at least as large as $E\Pi_C(K_T)$. Hence, $E\Pi_C(K_C) \geqq E\Pi_C(K_T) > \Pi_T(K_T)$, and the utility always prefers conventional contracting to the use of turnkey contracts.

Further, more units of capital are ordered under conventional contracting than in the turnkey situation. To see this, rewrite the first-order condition in the case of a conventional contract as

$$-\frac{w}{r} \frac{\partial L(P_0, K_C)}{\partial K} = \frac{\int_0^{p^0(K_C)} pf(p)\,dp}{F(p^0(K_C))} \equiv E(p \mid p \leqslant p^0(K_C))$$

---

[6] P_0 is assumed to be such that $\Pi_T(K_T) \geqslant 0$ for $K_T > 0$.

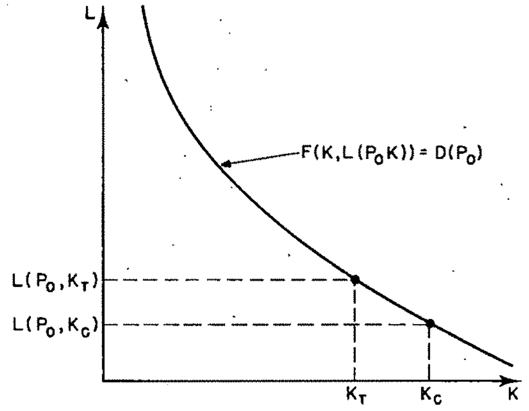


FIGURE 4. MARGINAL CONDITIONS FOR TURNKEY VS. CONVENTIONAL CONTRACTING

The right-hand side is the expected value of p given that $p \leqslant p^0(K_C)$. Except for the degenerate case $F(p^0(D_C)) = 1$, we have $E(p | p \leqslant p^0) < E(p) = \int_0^\infty pf(p)dp$. In the turnkey case, the first-order conditions reduce to

$$\frac{w}{r}\frac{F_K(K_T, L_T)}{F_L(K_T, L_T)} = E(p)$$

where $F(K_T, L_T) = D(P_0)$, hence $-\partial L(P_0, K_T)/\partial K = F_K(K_T, L_T)/F_L(K_T, L_T)$ so that we have $-(w/r)(\partial L(P_0, K_T)/\partial K) = E(p)$ for the turnkey case, hence $(-)(\partial L(P_0, K_C)/\partial K) < (-)(\partial L(P_0, K_T)/\partial K)$. As $(-)(\partial L(P_0, K)/\partial K)$ is the marginal rate of substitution between capital and variable factors along the isoquant $F[K, L(P_0, K)] = D(P_0)$, quasi concavity of F implies $(-)(\partial L(P_0, K_C)/\partial K) < (-)(\partial L(P_0, K_T)/\partial K)$ if and only if $K_C > K_T$ as shown in Figure 4. Thus we have established:

PROPOSITION 2: *In a Joskow-type model of utility-initiated regulatory review in which the rate-of-return constraint is never active and in which demand for output is inelastic over the entire range of outcomes, an expected income-maximizing utility will prefer a conventional contract to a turnkey contract* ($E\Pi_C > \Pi_T$), *and will order more units of capital under a conventional contract than under a turnkey contract.*

The intuitive reason for the utility's preference for a "risky" alternative to the risk-free turnkey contract is that the nonnegativity of profits constraint enables the utility to pass on to its customers cost overruns, while the utility itself captures excess profits from low observed values of p.

The assumption that demand is inelastic over the entire range of outcomes is unrealistic. However, the major result of Proposition 2, namely that utilities prefer conventional contracting to a turnkey contract, still holds when demand is elastic for sufficiently large P. To see this, let P_2 be such that $P_2 D(P_2)$ maximizes revenue over all values of P. Define $p^1(K)$ as that value of p such that

$$P_2 D(P_2) - wL(P_2, K) = rp^1(K)K$$

Thus, for a given value of K, $p^1(K)$ is the price per unit of capital that is so large that only if $P = P_2$ will the utility manage to break even; for any larger value of p, profits will be negative regardless of output prices. Clearly for $p > p^1(K)$, the utility will request that P be set at P_2 by the regulatory agency unless profits are actually higher in elastic range of demand. For simplicity, we will assume that Π is higher at P_2 than at $P > P_2$, but our conclusion in no way depends on this assumption.

Expected profits now are given by

$$\begin{aligned} E\Pi_C = \int_0^{p^0(K)} &[P_0 D_0(P_0) \\ &- wL(P_0, K) - rpK] f(p)dp \\ + \int_{p^0(K)}^{p^1(K)} &[P_1(p, K, L)D(P_1) \\ &- wL(P_1, K) - rpK] f(p)dp \\ + \int_{p^1(K)}^{\infty} &[P_2 D(P_2) - wL(P_2, K) \\ &- rpK] f(p)dp \end{aligned}$$

In this expression, the first integral is positive, the second is zero, and the third is negative.

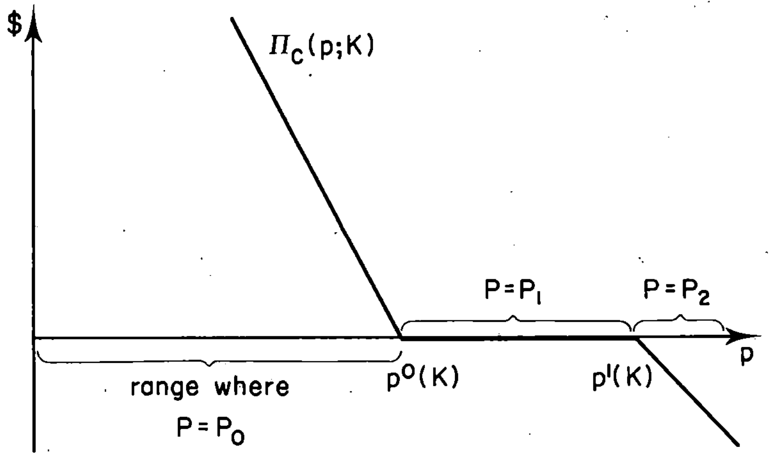


FIGURE 5. $\Pi_C(p;K)$ AS A FUNCTION OF p FOR GIVEN K

The first-order condition characterizing an optimum is then

$$\int_0^{p^0(K_C)}\left(-w\frac{\partial L(P_0,K_C)}{\partial K}-rp\right)f(p)\,dp$$

$$+\int_{p^1(K_C)}^{\infty}\left(-w\frac{\partial L(P_2,K_C)}{\partial K}-rp\right)f(p)\,dp=0$$

where K_C denotes the optimizing value of K. Then

$$E\Pi_C(K_C)=\int_0^{p^0(K_C)}\left[P_0D(P_0)-wL(P_0,K_C)-rpK_C\right]f(p)\,dp$$

$$+\int_{p^1(K_C)}^{\infty}\left[P_2D(P_2)-wL(P_2,K_C)-rpK_C\right]f(p)\,dp$$

But it is clear that the same argument applied earlier to show that $E\Pi_C(K_C)>\Pi_T(K_T)$ when demand was inelastic applies here as well, since in the range $p>p^1$, revenue is constant; thus the argument given earlier applies in full to this problem. Hence again the utility prefers the conventional contract to a turnkey contract.

Given that demand is elastic for $P>P_2$, will the utility continue to hire more units of capital under a conventional contract than under a turnkey contract? Rewrite the first-order condition above as

$$-\frac{w}{r}\left(F(p^0)\frac{\partial L(P_0,K_C)}{\partial K}+(1-F(p^1))\frac{\partial L(P_2,K_C)}{\partial K}\right)$$

$$=\int_0^{p^0(K_C)}pf(p)\,dp+\int_{p^1(K_C)}^{\infty}pf(p)\,dp$$

Because $\partial L(P,K_C)/\partial K$ is evaluated at both $P=P_0, P=P_2 (X=D(P_0), X=D(P_2))$, it is clear that there is no way to determine the relationship between K_T and K_C for an arbitrary quasi-concave production function $F(K,L)$. Hence we have:

PROPOSITION 3: *Under the conditions of Proposition 2, but dropping the assumption that demand is always inelastic, then an expected income-maximizing utility will prefer a conventional contract to a turnkey contract. If demand is always inelastic, more capital is ordered under conventional contracting than under a turnkey contract, but the comparison is ambiguous if elastic portions of the demand curve exist.*

Given the Joskow-type model of this section, the utility tends to prefer "risky" options to safe options because utility-initiated regulatory proceedings either eliminate or diminish the magnitude of negative profits. Figure 5 illustrates the profit function for a regulated utility under utility-initiated rate-of-return review. For a given value of K, profits $\Pi_C(p;K)$ are positive with $P=P_0$ for $p<p^0(K)$; $\Pi_C(p;K)=0$ with $P=P_1(p,K)$ for $p^0(K)\leqslant p\leqslant p^1(K)$; $\Pi_C(p,K)<0$ with $P=P_2$ for $p>p^1(K)$.

In fact, it can be shown that given a choice among probability distributions $f(p)$ all with the same mean $E(p)$, the Joskow-type utility operating with only the non-negativity of profits constraint will prefer that distribution with maximum dispersion. To show this, assume for simplicity that there is some finite $p=p_H$ such that $f(p)=0$ for $p>p_H$, and that profits $\Pi(p)$ are bounded at $\Pi(0)$. Formally, assuming demand is always inelastic, the utility is to choose $f(p)$ so as to

$$\max_{f(p)} \int_0^{p_H} \Pi(p)f(p)\,dp$$

subject to $f(p)\geqq 0, \quad \int_0^{p_H} f(p)\,dp=1,$

$$\int_0^{p_H} pf(p)\,dp=E(p)$$

Where $\Pi(p)\geqq 0$ for all p (demand is always inelastic) and $d\Pi/dp<0$ for $p<p^0$; $d\Pi/dp=0(\Pi(p)=0)$ for $p>p^0$. Then the optimum choice of $f(p)$ is

$$f(p)=\begin{cases} 1-\dfrac{E(p)}{p_H} & \text{for } p=0 \\ 0 & \text{for } 0<p<p_H \\ \dfrac{E(p)}{p_H} & \text{for } p=p_H \end{cases}$$

The rationale underlying this solution to a linear variational problem is perhaps seen most easily if one considers the discrete counterpart of the problem, where p takes on the values $p_0,\dots,p_N, p_i>p_j$, for $i>j, p_0=0, p_N=p_H$. Let $x_i=\{$probability with which $p=p_i\}$. Then the problem becomes

$$\max_{\{x_i\}} \sum_{i=0}^{N} \Pi(p_i)x_i$$

subject to

$$x_i\geqq 0, \quad \sum_{i=0}^{N} x_i=1, \quad \sum_{i=0}^{N} p_i x_i=E(p)$$

This LP problem involves $N+1$ variables and two constraints (excluding nonnegativity constraints), hence an optimal solution need involve no more than two positive x_is. Since $\Pi(p)$ is decreasing and convex in p, as much weight as possible should be placed on $\Pi(0)$ subject to the expected value constraint. Thus x_0 attains its largest admissible value when $x_i=0$, $i=1,\dots,N-1$ and $x_N=E(p)/p_H$, with $x_0=1-E(p)/p_H$.

Note that this conclusion extends to the case of elastic portions of the demand function as well, except that p_H now becomes the maximum value of p such that $\Pi(p)=0$.

PROPOSITION 4: *In a Joskow-type model as in this section, given a choice among probability distributions over the price per unit of capital all with the same expected value* $E(p)$, *the utility will choose that distribution with maximum dispersion over values of* p *that produce nonnegative profits, with optimal choice* $Pr\{p=0\}=1-E(p)/p_H$, $Pr\{p=p_H\}=E(p)/p_H$, *where* p_H *is the largest value of* p *such that* $\Pi(p)=0$.

A feature of the Joskow argument which has been ignored thus far is the fact that regulatory proceedings take time and are not costless. During proceedings cost overruns must be borne by the utilities. One would expect that cognizance of these costs would weaken or reverse some of the results derived above, but this is not necessarily the case. Although the effects of cost overruns and regulatory review involve dynamic considerations, we stylize them within the context of our simple static model so as to maintain symmetry with the earlier exposition.

To incorporate these possibilities we let $A(p)$ be the annualized costs of the regulatory proceedings and $C(p)$ be the annualized costs of cost overruns. We define $p_L(K)$

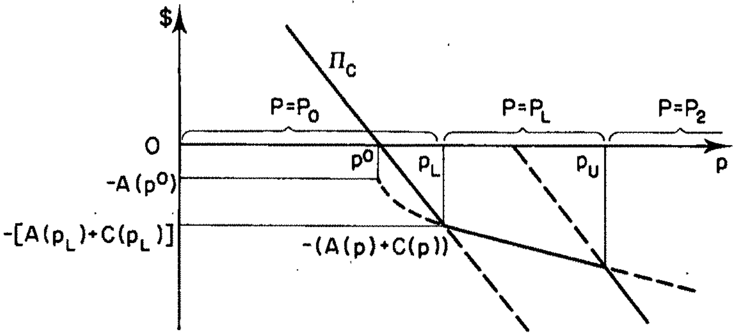


FIGURE 6. $\Pi_C(p,K)$ WHEN REGULATORY PROCEEDINGS ARE NOT COSTLESS

as the lowest[7] price of capital for which

$$P_0D(P_0) - wL(P_0,K) - rp_L(K)K \equiv -[A(p_L)+C(p_L)]$$

That is, $p_L(K)$ is the lowest price of capital for which it is marginally worthwhile for the utility to initiate review proceedings. While $A(p)>0$ for all p, $C(p)=0$ for $p \leqq p^0(K)$: however $p_L(K)>p^0(K)$ so that $C(p_L)>0$. Also it is natural to assume that $A(p)$ and $C(p)$ are nondecreasing functions of p. When $p>p_L(K)$ and p is close to p_L the utility initiates review proceedings and rates are hiked to P_L, where

$$P_LD(P_L) - wL(P_L,K) - rp_L(K)K = 0$$

Clearly $P_L > P_1$, as we explicitly assume that the costs of the review proceedings and cost overruns must be borne by the utility and cannot be passed on to consumers.[8]

In a similar fashion we define $P_U(K)$ by

$$P_2D(P_2) - wL(P_2,K) - rp_U(K)K = -[A(p_U)+C(p_U)]$$

assuming again for expositional simplicity that $p_U(K)$ is unique for each K and that

$$P_2D(P_2) - wL(P_2,K) - rpK < -[A(p)+C(p)]$$

for $p > p_U(K)$, where P_2 is defined as before. We now have

$$\Pi_C = \begin{cases} P_0D(P_0) - wL(P_0,K) - rpK & \text{for } p \leqslant p_L(K) \\ -A(p) - C(p) & \text{for } p_L(K) \leqslant p \leqslant p_U(K) \\ P_2D(P_2) - wL(P_2,K) - rpK \\ \quad -A(p) - C(p) & \text{for } p \geqslant p_U(K) \end{cases}$$

By the structure thus imposed we have $\Pi_C(p,K) \geqslant \Pi_T(p,K)$ for all p with strict inequality for $p > p_L(K)$. By the order-preserving property of the integral, $E\Pi_C(p,K) \geqslant \Pi_T(E(p),K)$, with strict inequality if $Pr\{p > p_L(K)\} > 0$; likewise, similar to the earlier demonstration, $K_C > K_T$ provided demand is inelastic. We summarize these results as Proposition 5 and depict the profit function in Figure 6.

PROPOSITION 5: *Under the conditions of Proposition* 3 *where, in addition, the utility incurs costs* $A(p)$ *upon initiating rate review proceedings and* $C(p)$ *from cost overrun during proceedings, conventional contracting is preferred to turnkey contracting. Moreover, if demand is inelastic over all possible outcomes, more capital is ordered under a conventional contract than the turnkey contract.*

---

[7] For simplicity we suppose that for each K, $p_L(K)$ is unique and that $P_0D(P_0) - wL(P_0,K) - rpK < -R(p) - C(p)$ for $p > p_L(K)$. However, our argument generalizes (with some notational clutter) for multiple $p_L(K)$.

[8] If this were not true then $p_L = p^0$ and P_L would be defined by $P_LD(P_L) - wL(P_L,K) - rp^0(K)K = -A(p^0)$ as $C(p^0)=0$. However, this does not qualitatively affect the analysis.

In effect, this result states that whenever initiating a rate review is preferred to foregoing the review, the utility with the conventional contracting avails itself of this option and improves its position vis-à-vis the turnkey utility. While the possibility of multiple $p_L(K)$s and $P_U(K)$s complicates the analysis, it does not alter the conclusion.

III. The Turnkey Era in Nuclear Power

An application of the theory developed in this paper is to the so-called turnkey era in the development of the commercial nuclear power industry in the United States, roughly from 1963 through mid-1966. Prior to this period, nuclear power was a heavily subsidized stepchild of the Atomic Energy Commission, which assisted in research and development activities of reactor manufacturers and utilities. But in 1963, first General Electric and then Westinghouse—the two leading reactor manufacturers—announced the availability of fixed-price contracts for complete nuclear power units in the 600–1,000 MWe (megawatts electric) range, at prices that made nuclear power competitive with coal for much of the country. These were called turnkey contracts because the reactor manufacturers agreed to take on all of the responsibilities for design, building, and testing of the units (including meeting regulatory guidelines), simply turning the key over to the utility when the plant was ready for commercial operation. Moreover, these turnkey contracts were available at guaranteed prices.

The turnkey era ended in June of 1966 when General Electric announced the termination of its turnkey contract program. During the era there were twenty nuclear units ordered, thirteen under fixed-price contracts and seven under conventional (cost-plus) contracts with utilities bearing cost risks in the case of the conventional contracts.

The most striking features of the turnkey era were first the fact that actual costs (to the reactor manufactureres) of the turnkey units far exceeded quoted prices, so that General Electric and Westinghouse lost a combined total of nearly $1 billion on the thirteen turnkey units; and second, that the close of the turnkey era was followed by an upsurge in orders for nuclear units. Thus orders for nuclear units were four in 1963, zero in 1964, seven in 1965 and eight in 1966 through June; twelve in the second half of 1966, and thirty in 1967.

There are of course a number of explanations for the rise of the nuclear power industry from the mid-1960's on: increases in capital and fuel costs of coal, natural gas, and oil units, demonstration of the technical feasibility and economic viability of large nuclear units, the rise of environmental concern with stack emissions and the like. The theory developed in this paper argues that another factor at work was the regulatory environment that utilities found themselves in from the 1960's on; namely an environment in which the operative constraint was the zero-profit constraint rather than the allowed rate-of-return constraint.

When an attempt is made to test the theory at the level of individual utilities placing orders for nuclear units, the results are ambiguous. The theory argues that utilities bumping against the rate-of-return constraint should find fixed-price units desirable, while utilities at the zero-profit constraint should find conventional units the better choice. But of course there are problems in moving from the simple model presented above to decision problems facing utilities, not the least of which is the long lead time between ordering a nuclear unit and the time it comes on line, much less the time before it is retired. And certainly it is the utility's judgment of its prospective regulatory situation over such future periods that is relevant in making choices as to contracting for nuclear units.

Beyond this, the fact that the turnkey units ended up costing from two to three times the price quoted by General Electric or Westinghouse means that they turned out to be bargains relative to the nonturnkey units. Differences in beliefs as to the probability distribution of costs provide more than enough of a rationale for choices between turnkey and nonturnkey contracts.

Table 1 presents data on actual rates of return on rate-base capital together with

TABLE 1—RATES OF RETURN ON RATE-BASE CAPITAL AND ALLOWED RATES OF RETURN 1961–65, 1966–70, BY UTILITIES ORDERING NUCLEAR UNITS DURING THE TURNKEY ERA

Company and Plant and Ownership Share		Year of Order	Average Rate of Return[a] on Rate-Base Capital 1961–65	1966–70	Allowed Rate of Return[a] Pre-Turnkey 1957–65	Post-Turnkey 1966–70
Turnkey Units[c]						
Iowa-Illinois G&E						
Quad Cities 1&2	(25%)	1966	8.22	8.42	5.85 ('60)	–
Commonwealth Edison						
Quad Cities 1&2	(75%)	1966	8.00	7.95	5.85 ('58)	7.0 ('70)
Dresden 2&3		1965				
Northern States Power						
Monticello		1966	7.95	8.13	5.96 ('57)	–
Wisconsin Elec. Power						
Point Beach 1&2	(50%)	1966	6.49	6.81	6.0 ('58)	–
Wisconsin Mich. Power						
Point Beach 1&2	(50%)	1967	5.94	6.30	6.15 ('61) 6.0 ('64)	7.67 ('69) 7.65 ('70)
Carolina P&L						
Robinson 2		1966	7.14	7.02	4.08 ('63)[b]	–
Rochester G&E						
Ginna		1965	6.78	7.72	6.2–6.3 ('63)	6.2–6.3 ('68)
Jersey Central Power						
Oyster Creek		1963	6.87	7.38	6.0 ('60)[b]	7.57–7.75 ('70)
So. Cal. Edison						
San Onofre	(80%)	1963	6.64	6.88	6.35 ('61)	7.35 ('69)
San Diego G&E						
San Onofre	(20%)	1963	6.32	7.04	6.35 ('61)	7.35 ('69)[b]
Hartford Elec. Lt.						
Millstone	(28%)	1965	6.54	7.30	6.06 ('61)	7.5 ('70)
West Mass. Elec.						
Millstone	(19%)	1965	7.34	7.64	–	–
Conn. Lt. & Power						
Millstone	(53%)	1965	6.79	7.69	6.06 ('61)[b]	7.5 ('70)[b]
Connecticut Yankee	(25%)	1963				
Nonturnkey Units[c]						
Boston Edison						
Pilgrim		1965	7.06	7.69	5.7 ('59)	–
Niagara-Mohawk						
Nine Mile Point		1963	5.90	6.04	6.2–6.3 ('63)[b]	6.2–6.3 ('68)[b]
Florida P&L						
Turkey Point 3		1965	8.26	8.46	6.98 ('57)	6.95 ('66) 7.08 ('67) 7.1–7.6 ('69)[b]
Consumers Power						
Palisades		1966	7.20	7.64	6.36 ('63)	7.2 ('70)
Consolidated Edison						
Indian Point 2		1965	5.29	5.62	6.2–6.3 ('61)	6.2–6.3 ('68) 7.28 ('70)

Sources: Statistics of Privately Owned Public Utilities, 1961–70; Public Utility Reports.

[a] Shown in percent.

[b] Allowed rate of return for another utility in the state.

[c] Only utilities owning at least 10 percent of a nuclear unit are included. Indian Point 2 is classified as a turnkey unit in several other studies of the turnkey era. However, the unit was clearly not contracted for on a fixed price contract, as is indicated by an almost doubling of the reported price of the unit between the ordering date and the date of operation (from $121/kw to $212/kw). Hence we have classified Indian Point 2 as nonturnkey for purposes of this table. Browns Ferry 1&2 (TVA) was excluded because TVA is publicly owned.

allowed rates of return, for privately owned utilities ordering reactors during the turnkey era. All privately owned firms with an interest of 10 percent or more in a reactor ordered during the era are included in the table. Actual rates of return are those calculated by the Federal Power Commission, averaged over the five-year periods, 1961–65 and 1966–70. Allowed rates of return are those laid down in *PUC* and/or court decisions as of the dates indicated. Allowed rates of return vary over time and are stated formally only in the course of rate reviews which, as Joskow has pointed out, occurred relatively infrequently during the 1960's. As a consequence we have included all reported allowed rates of return (together with dates) for the utilities, between 1957 and 1970. For certain utilities, there were no rate reviews during this period so we have included in the table the allowed rate of return for other utilities in the same state.

We present data for the "pre-turnkey" (up to 1965) and "post-turnkey" (1966–70) eras, both for average rate of return and allowed rates of return. This permits one to use either an historical or a rational expectations approach to the issue of whether the actual rate of return is above or below the allowed rate of return.

Some simplistic comparisons can be made from the data in Table 1. Crudely, one might argue that high rates of return either during the ordering period (1961–65) or during the construction period (1966–70) would indicate a higher probability of bumping into the allowed rate-of-return constraint and conversely for low rates of return. But in fact there is no significant difference between the average rank of actual rates of return for turnkey and nonturnkey purchasers (average rank, turnkey purchasers $\simeq 8$, average rank, nonturnkey purchasers $\simeq 10$).

But because of differences in allowed rates of return among states, a ranking simply in terms of actual rates of return is certainly flawed. Comparing actual rates of return with allowed rates of return, we find that ten of twelve turnkey purchasers earned above the allowed rate of return in the 1961–65 period; and three of five nonturnkey purchasers earned above the allowed rates of return. In the 1966–70 period, two of eight turnkey purchasers earned above the allowed rate of return along with one of four nonturnkey purchasers.

While these crude comparisons tend slightly to agree with the theoretical conclusion that firms bumping against the allowed rate of return tend to favor turnkey contracts while firms bumping against the zero-profit constraint tend to favor conventional (nonturnkey) contracts, the empirical case for the theory has certainly not been established.[9] Perhaps other instances can be identified in the history of regulated utilities that provide less ambiguous results.

---

[9] Ron Braeutigam has noted that the subjective probability distributions over costs might have been dispersed enough so that a two-constraint model of the utility is needed to explain their choices.

REFERENCES

H. Averch, and L. L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

H. S. Burness, W. D. Montgomery, and J. P. Quirk, "The Turnkey Era in Nuclear Power," *Land Econ.*, May 1980, *56*.

P. Joskow, "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Regulation," *J. Law Econ.*, Oct. 1974, *17*, 291–327.

Federal Power Commission, *Statistics of Privately Owned Public Utilities*, Washington 1961–70.

Public Utility Reports, 3rd ed., Vols 40–100, 1961–70.

Do Union Members Receive Compensating Wage Differentials?

By GREG J. DUNCAN AND FRANK P. STAFFORD*

Empirical work based on sample survey data has shown large wage differentials between union and nonunion workers with similar measured personal skills. Several authors have hypothesized that these differentials are greater than might be explained by unionization or its underlying causes.[1] In this paper we offer two structural hypotheses which include a relation between unionism and conditions of work. Both hypotheses imply that part or all of the observed differential will be a compensating wage differential for certain conditions of work—notably a structured work setting, inflexibility of hours, employer-set overtime, and a faster work pace—rather than rents to union members or differentials attributable to unmeasured personal skills acquired on the job. While one hypothesis specifies unionism to be a function of these kinds of working conditions, the other specifies these working conditions to be an employer response to unionism.

Although we are unable to distinguish between these two hypotheses, we are able to establish the presence of compensating wage differentials for union members. Union members have a work setting with less personal choice in the way in which work is organized. For our sample of blue-collar workers we estimate that about two-fifths of the union-nonunion differential is a compensating wage differential. The research exploits unique aspects of several data sets.

I. Theoretical Approaches to Workplace, Public Goods, and Unionism

Work pace and work scheduling are different from the more widely researched demand for and supply of specific on-the-job training, and can be viewed as *substitutes* for highly specialized individual skills. Production organized on a process basis can break larger tasks into smaller components which require less individual skill. These simpler tasks must be sequenced and coordinated, and this will require greater economic incentives in the form of higher wages to compensate workers for reduced opportunities to choose an individual or flexible work schedule and an individual work pace. These production processes involve greater interdependencies among workers and between workers, and a larger capital stock per worker—the assembly line is a prototype of such a process. All workers in such a work group or plant will have a set of common arguments in their utility functions—work pace, length of workday, level of protection from injury and illness, temperature, and other conditions of varying importance.[2] Production may be greatly increased through a work period with specific beginning and ending points identical for all workers, a work period with few opportunities to take unscheduled breaks, and a specified work pace. Many of these working conditions are public goods because their consumption is nonrival. Consider, for example, the temperature in a large plant.

---

*University of Michigan. Research for this paper was supported by grants from the National Science Foundation (NSF) Special Projects Division and from NSF Office of National R and D Assessment to the Survey Research Center. The research reflects our views rather than those of any federal agency. An earlier version of this paper appeared under the title, "Pace of Work, Unions and Earnings in Blue Collar Jobs." We would like to thank Theodore Bergstrom, George Borts, Richard Freeman, Daniel Hamermesh, F. Thomas Juster, James Medoff, Jacob Mincer, Melvin Reder, Daniel Saks, and Gary Saxonhouse for comments. Dorothy Kempter provided able research support.

[1] For a review of this literature, see George Johnson.

[2] Length of the workday can be common to all workers in a given work group (see Alan Deardorff and Stafford).

While a "comfortable" temperature can be provided by the firm to its workers at some cost, any given temperature level is consumed in equal amounts by all workers. Other working conditions can be similarly shared by all of the workers in the plant. These public good aspects of the workplace imply restrictions on individual behavior or an imperfect matching of an individual worker's preferences with plant policy. Such shared working conditions raise questions about possible mechanisms, such as labor unions, for choosing levels of the goods and compensation of the workers.

The public goods potential of the work place can also arise from the nature of the input of effort measured, say, by hours or pace of work. It is common to assume separability of the effort input of each worker, but there are many cases when the input of each worker may be highly complementary. Suppose output Q is produced according to

$$Q=Q(E,K) \tag{1}$$

where E is total effort of the workers and K is some other input. Then at one extreme we can express E as

$$E=\sum_{i=1}^{n} \alpha_i e_i \tag{2}$$

where e_i is per worker effort (hours of work, for example), α_i is an adjustment for efficiency differences across workers, and n is the number of workers. Alternatively, suppose

$$E=\min\left(\frac{e_1}{\alpha_1},\frac{e_2}{\alpha_2},\dots,\frac{e_n}{\alpha_n}\right)n \tag{3}$$

Then if $\alpha_1=\alpha_2=\dots=\alpha_n$ all workers should work at the same level of effort, for the lowest level of effort by any one worker will limit the labor input, and a single zealous worker will not add to output. If (2) characterizes a typing pool, then (3) can be thought of as characterizing an assembly line. In (3) one worker could unilaterally buy additional leisure, perhaps in the form of an unscheduled break, by paying the full value of reduced output $(p_1(\partial Q/\partial E)(\partial E/\partial e_i)$, where p_1 is output price), but others could also realize added leisure without the firm incurring loss of additional output just as in the usual public goods case where "free riders" may receive benefits.

There is also the question of group formation involving search costs. A group with more homogeneous preferences will make the public goods problem implied by (3) less difficult but will be more costly to organize. If the costs of search per worker are high, the resulting heterogeneity will add a nontrivial dimension to the public goods problem. The situation is complicated by the fact that workers who are dissatisfied with the workgroup's policies will leave; hence economizing on search costs will require higher wages to meet the transfer utility of some workers, particularly if they are free to leave for sectors characterized by (2).

Before turning to a specific model it is useful to summarize the more general applicability of local public finance theory for the organization of groups. In the context of local public finance, Charles Tiebout provided an important insight by pointing out that the public good aspects of an economic unit need not involve a collective choice mechanism if people can choose among alternative units which offer different combinations of public goods.[3] If, for example, an individual is dissatisfied with the services provided by a local government, he can simply move somewhere else. At the other extreme, much of the public finance literature has centered on possible mechanisms for the provision of public goods under the assumption of no mobility (see Theodore Groves and John Ledyard; Edward Clarke; Howard Bowen). The reason that a blend of these two approaches may be fruitful in analyzing the public goods aspects of the workplace is that nonzero search and mobility costs prevent a full Tiebout solution. On the other hand, these costs are not usually so high as to result in a pure public goods solution; that is, a zero mobility equilibrium implied in the Samuelson model (see Paul Samuelson).

---

[3] Had Tiebout written his paper after Sherwin Rosen (1974) he might have said that citizen-consumers choose a community and buy a set of characteristics and that these characteristics have implicit (hedonic) prices.

There is a tradeoff involving the costs of using the market versus the costs of using some mechanism for public choice inside the economic unit, particularly in the short and intermediate run, when there will be incomplete adjustment to exogenous shocks experienced by the unit. The relative use of the market versus internal mechanisms will therefore vary across different types of units in real markets. While it is possible to organize a contingent to see the manager about the customer policies of a given fast food place, it is generally more effective to choose an alternative place to eat, but if one is displeased by national tax policy, one may opt for writing one's senator or joining an interest group rather than emigrating.

For the public goods aspect of the workplace to be relevant two conditions have to be met in addition to complementarities in consumption or work effort. First, there must be substantial costs of using the market in the form of search costs for the firm in finding new workers and mobility costs for workers who quit. Second, the production unit needs to be subject to a number of imperfectly foreseen, exogenous shocks over time—changes in product market, technology, and the like—which will require adjustments to meet efficiency conditions. In light of the new economic environment, the firm can be viewed as inheriting a stock of workers, capital, etc. To economize on search costs, the inherited workers should be canvassed as to preferences with respect to the acceptability of potential responses to the exogenous events in terms of longer or shorter workdays, different work assignments, pace of work, and so on. This internal demand-revealing function can be hypothesized as a purpose for a union, relating to the general reference to "industrial democracy" and an emphasis on shared working conditions in collective bargaining.

Suppose the firm discovers a new technology which makes hours of work more productive and it would like to have its workforce supply more hours of work per worker. If search were costless, the firm might decide on a strategy of seeking out workers with preferences for longer work hours (and perhaps paying a wage premium) and firing its stock of inherited workers (who would require an even larger premium to work longer hours). Note, of course, that if there were no interdependencies or complementarities between hours worked of the different workers (i.e., as in (2) where total labor hours was the relevant argument in the production function), the firm could get more labor hours by adding more workers and by keeping the hours of the inherited workers at the preshock levels. Adjustment via the market or internal adjustments is far less important with such technology, and there would be less need for collective decision making about an appropriate response by the workers.

Let us illustrate formally the efficiency conditions in light of search costs and the potential role of demand revealing mechanisms inside the firm when production requires the joint presence during the workday of all workers in the firm. Consider a firm which seeks to maximize profits (Π),

$$\Pi = p_1 Q - W_A - W_M - cS \tag{4}$$

where Q is output sold at price p_1, W_A is the wage bill paid to the A searched workers who are accepted for employment (accepted workers), W_M is the wage bill paid to the M of the I inherited workers who are retained (retained workers), c is the unit search cost,[4] and S is the number of prospective employees searched. The firm's production function is given as

$$Q = Q(h, A+M) \qquad A \leqslant S, M \leqslant I \tag{5}$$

It is not usual to specify hours per worker (h), and number of workers ($A+M$) as separate arguments in a production function. Depending on the capital associated with the production process, the firm will not want an excessive number of workers at a given location (the size of the building interior provides one constraint!), and the productivity of work hours per worker certainly declines beyond some point.[5] Here the source of the potential public goods aspect is the fact that the hours per worker input is written as

---

[4] The search costs can also include trial period training costs and in general can represent costs of bringing new employees into the firm.

$$\min\left(\frac{h_1}{\alpha},\frac{h_2}{\alpha},\ldots,\frac{h_M}{\alpha},\ldots,\frac{h_{A+M}}{\alpha}\right)$$

Since there is no added output to any worker who works "after hours," $h_1=h_2=\ldots=h_M=\ldots=h_{A+M}$ is replaced in (5) with a single decision variable, h.

The inherited workers have a frequency distribution of preferences for hours of work given by the parameter θ with density $f(\theta)$. The firm is a distribution taker in searching for new workers from the market with a density function given by $p(\theta)$. The firm meets the going utility to attract the workers it accepts[6] and knows that reservation utility $\overline{U}_i$ is given by

$$(6)\qquad \overline{U}_i=g(\theta)$$

$$(7)\qquad U_i=U_i(X,h;\theta)$$

for all workers regardless of whether they are inherited or found.

The average wage bills for the accepted and retained workers are

$$(8)\qquad W_A=S\int_{\theta'_Z}^{\theta'_U}p(\theta)X(\theta,h)d\theta$$

$$(9)\qquad W_M=I\int_{\theta'_z}^{\theta'_u}f(\theta)X(\theta,h)d\theta$$

where $X(\theta,h)$ is the *daily* wage that the firm must pay to meet transfer utility (6) of each worker, and upper and lower bounds for searched and inherited workers are (θ'_U,θ'_Z) and (θ'_u,θ'_z), respectively. Here we assume the firm does not engage in sequential sampling but searches for its workforce and reaches its employment and hours decision by a search "all at once" with its employment agents given a decision rule to accept searched workers who fall in the $(\theta'_U-\theta'_Z)$ interval and to pay the accepted workers enough to meet their transfer utility.

Optimization consists of maximizing the expectation of (4) subject to (5)–(9) via the firm's decision variables θ'_U, θ'_Z, θ'_u, θ'_z, S, and h. To simplify the exposition suppose that $p(\theta)$ and $f(\theta)$ are rectangular with upper and lower limits of (θ_U,θ_Z) and (θ_u,θ_z), respectively, and that the utility functions of the workers are of a "separable" form, commonly used in demand-revealing models, specifically,

$$(10)\qquad U_i(X_i,T-h_i)=X_i+\theta_i ln(T-h_i)$$

with T the total hours available per day. Let (6) be of the form

$$(11)\qquad \overline{U}_i=V+b\theta_i \qquad (b<0,\theta_i>0)$$

The price of the market or private good is normalized at 1.

Let $R=A+M$. Then first-order conditions for an interior optimum are[7]

$$(12)\qquad p_1Q_R=V+\frac{(b-ln(T-h))(\theta'_U+\theta'_Z)}{2}+c\left(\frac{\theta_U-\theta_Z}{\theta'_U-\theta'_Z}\right)$$

$$(13)\qquad p_1Q_R=V+\frac{(b-ln(T-h))(\theta'_u+\theta'_z)}{2}$$

$$(14)\qquad V+\frac{(b-ln(T-h))(\theta'_u+\theta'_z)}{2}=V+\frac{(b-ln(T-h))(\theta'_U+\theta'_Z)}{2}+c\left(\frac{\theta_U-\theta_Z}{\theta'_U-\theta'_Z}\right)$$

$$(15)\qquad p_1Q_h=\frac{1}{2(T-h)}\left[\frac{S(\theta'^2_U-\theta'^2_Z)}{(\theta_U-\theta_Z)}+\frac{I(\theta'^2_u-\theta'^2_z)}{(\theta_u-\theta_z)}\right]$$

---

[5] For empirical evidence on the demand for hours per worker and number of workers, see Frank Brechling.

[6] Subsequent first-order conditions can be easily modified for the case where the firm must set a single hours-earnings package and cannot pay exact transfer utility.

[7] Appropriate Kuhn-Tucker conditions are necessary when $\theta'_z=\theta_z$ or $\theta'_u=\theta_u$.

From (12) we can say that given an hours decision and the upper and lower bounds for θ on searched workers, the firm should carry search to the point where the marginal revenue product of an additional worker, $p_1 Q_R$, equals the expected daily wage $V+(1/2)(b-ln(T-h))(\theta'_U+\theta'_Z)$, plus the expected search costs per worker accepted, $c((\theta_U-\theta_Z)/(\theta'_U-\theta'_Z))$. The expected wage costs are, under the assumption of a rectangular probability density function, a simple average of the transfer daily wage of workers at the lower and upper bound acceptance values of θ. Expected search costs per worker increase as the acceptance interval $\theta'_U-\theta'_Z$ narrows. From (13) we have an analogous first-order condition for retained workers except for expected search costs. Since retained and accepted workers are specified as perfect substitutes in production, the resulting equality between (12) and (13) directly implies (14). From (14) we can see that for interior solutions there will be a narrower acceptance interval set for searched workers than for inherited workers, and this acceptance interval difference should be greater the greater the search costs. Knowing this, the inherited workers (or some coalition of them) may want to capture some of their inertial rents—windfalls which arise because they are there and can be replaced only with costly search.

Given the upper and lower acceptance bounds and a search decision, (15) implies that hours be set so that the marginal revenue product of hours per worker be equated to marginal wage costs of hours. Marginal wage costs of hours rise with increasing hours and are an increasing function of the disparity between the upper and lower acceptance bounds, $(\theta'^2_U-\theta'^2_Z)$ and $(\theta'^2_u-\theta'^2_z)$. Rewrite (15) as

$$(16)\quad p_1 Q_h=\frac{A(\theta'_U+\theta'_Z)/2}{T-h}$$

$$+\frac{M(\theta'_u+\theta'_z)/2}{T-h}\qquad (A\leqslant S, M\leqslant I)$$

It can be seen as a particularization of the well-known condition set out by Samuelson for efficiency in the provision of public goods. In the context of the current model it says that given the number of workers and the level of heterogeneity of preferences among retained and accepted workers, work hours per worker should be chosen by the firm so that the ratio of leisure's price (the foregone marginal revenue product of additional hours per worker) to the price of the private good (normalized at 1) is equated with the sum of the marginal rates of substitution of leisure (the public good) for the private good.

The model emphasizes the public good aspects of particular working conditions and the role of search or firm-specific training costs in conditioning the heterogeneity of the workgroup for which the public goods problem is to be solved. The role of a mechanism to assess and voice preferences of existing workers is more apparent if we relax the convenient functional form on utility. Then from any previous equilibrium the firm will inherit people with different marginal willingness to accept compensation for *changed* hours (or other *shared* working conditions). In this case their acceptance of a previous (X,h) combination does not provide complete information about their willingness to accept alternative (X,h) combinations and implies an economic value to a mechanism for assessing this.[8] Institutionally, such a mechanism could be worker councils or meetings of union locals. In an academic setting it could consist of the department chairperson and executive committee members canvassing their colleagues about workplace issues of common interest.

Although equations (12)–(15) characterize efficient solutions for the common working condition, they do not determine short-run distribution. Suppose that capital will leave or its future availability is jeopardized if not given sufficient (zero) profits. Then workers would want to maximize utility subject to a minimum profit constraint. Conversely, if future hires will be jeopardized, or if workers will leave if not given reservation utility, then the owners of capital will want to maximize profits subject to reservation utilities ($\bar{U}_i$) of the workers. In each case efficiency requires (12)–(15), but the distribution of the profits between workers and capital owners is not determined by these efficiency conditions. As a result short-run redistribution is a potential source of dispute. Note also that whether workers are "in control" or whether we have a "company union," there will be incentives to run some type of demand-revealing mechanism (median voter, Clarke tax) to maximize the size of the pie to be redistributed. That is, if workers are in control they will want to avoid excessive mobility since this will imply search costs that will cut into their residual share of the pie. In general, where search and mobility are a part of the optimum, they may want to bribe coworkers to leave and hire some new coworkers, but will also want an effective demand-revealing mechanism to reach the decision.[9]

---

[8] If it is known beforehand that there will be variations in the firm's demand for hours per worker, the firm would be searching for individuals willing to accept different hours at not drastically different compensation per hour. This is the utility analogue of George Stigler's production function with adaptability.

A complete analysis should include search decisions in light of a market distribution of θ among potential employees (with perhaps a fixed relation between θ and $\bar{U}$ with $\partial\bar{U}/\partial\theta > 0$).[10] If θ is viewed as given by a continuous probability density function, then no firm will ever be able to find perfectly identical workers without bankrupting itself. Hence, even with a stationary economic environment in equilibrium, there will be a dispersion of θ in the firm's workforce. Consequently, even with perfect wage discrimination (i.e., personally tailored wages to meet reservation utility of each worker), the firm will have to offer wage bribes to some of the work lovers and work haters which it employs or, for example, to employees differing in relative preference for safety, assuming there is an indivisibility in the provision of safety to a group of workers. The implication is that, given skill (market alternatives), workers in work environments that are structured via complementarity of effort will receive compensating wage differentials even if the demand revelation process is functioning perfectly. But these differentials are smaller than they would have been without the demand-revealing mechanism.

We hypothesize that the production settings that give rise to interdependencies or complementarities among workers and to the demand for work requirements by the employer to coordinate worker effort also give rise to labor unions. Unions, especially industrial unions, are most likely to be found in these production settings in part because organizing costs are low and because unions can represent a mechanism for dealing with the public good problems within the workplace. In this latter view, industrial unions represent the labor component of the firm, which Ronald Coase has argued represents a collection of formal or informal rules for resource allocation. As such, a union can embody mechanisms or rules which are substitutes for market transactions which organize activities across firms; that is, a union can be a mechanism for economizing on search and mobility costs.

We also postulate that a capital-intensive production technology typically implies a payoff to coordinated worker effort (a complementarity of effort as in equations (3) or (5)). Further, if the capital stock depreciates at a rate independent of the daily use rate (as with obsolescence), there will be incentives to utilize the capital on a twenty-four-hour basis. To do so will require compensating wage differentials to encourage workers to work at a pace and/or schedules consistent with a fuller use of the capital stock (for example, shifts). This can be called the capital stock utilization hypothesis and is set forth in a paper by Deardorff and Stafford. In summary, we hypothesize that unions will be more prevalent in capital-intensive continuous-process activities which give rise to worker effort complementarities, corresponding public goods problems, and incentives to schedule work in a way that will more fully utilize the capital stock.

---

[9] This suggests, of course, that the problem is not simply distribution between workers and management but among workers as well. A coalition of workers may want to maximize their utility and this may entail their "firing" the noncoalition members.

[10] That is, the market may generate a higher "going utility" for work lovers and there is substantial empirical evidence of higher wage rates for longer hours.

By our emphasis in this paper on organization of work and its relation to unionism, we do not wish to imply that other reasons for unions (political, occupational licensure, etc.) are not important, nor do we wish to imply that variables such as geographic concentration in labor markets, which Harold Levinson has argued lowers the cost of organizing, should be overlooked. We believe that unions are quite heterogeneous with respect to purpose and conditions which support them. However, our data restrict us to analyzing the relation between measures of the workplace and personal characteristics since we do not have measures of marketwide characteristics.

If there is validity to the hypothesis that certain work settings give rise to unions as a mechanism to deal with public goods, and that these work settings imply reduced personal choice even with the mechanism functioning, then there will be a relation between these working conditions and unionism. The higher earnings of union members will reflect, in part, compensating wage differentials rather than rents. Consequently, working condition variables may constitute a set of variables omitted from conventional analysis of the determinants of union membership and effect of union membership on wages and, if so, the estimated union-nonunion wage differential may overstate the causal relationship. For use below we shall call the hypothesis that shared levels of work effort and working conditions give rise to the need for compensating wage differentials and unions the interdependencies hypothesis.

An alternative, more traditional explanation for union wage effects is entry barriers which may cause, at least initially, large wage differences with unionized firms operating further back on the factor-demand schedule. This can be modified into a second hypothesis which we call the union-partial-control hypothesis. This hypothesis emphasizes that if the union controls only selected aspects of employment, the firm can respond by varying other decision variables. This is illustrated by the quality adjustment case advanced by H. Gregg Lewis. Employers endeavor to cope with the high union wages by using the high wages to attract better quality workers, and eventually the initial earnings advantage of the workers is eroded.[11] Though there can be inefficiencies from employment of workers who are "too skilled" for the job, in the limiting case, with sufficient turnover, the workers will realize (net) earnings only equal to those in their best alternative employment.[12]

Other variables which the firm could change in response to the higher wage include the capital-labor ratio, working conditions, or pace of work.[13] Unless timing of response can be ascertained, the cross-sectional observation of high wages, larger amounts of capital per worker, and a more structured work environment would be consistent with both the interdependencies hypothesis and the partial control hypothesis. In summary then, both hypotheses are consistent with poorer working conditions for union members combined with a wage advantage, at least part of which will be a compensating wage differential.

II. Wage and Membership Equations

In this section, the various hypotheses about working conditions, unionization, and wage rates are formalized into a series of wage and union membership equations. A convenient starting point for our discussion is the widely used human capital-labor quality wage equation modified to account for union membership and working conditions and a second relation specifying union membership as a function of working conditions and other variables:

---

[11] As Lewis notes: "For example, a 25 percent greater relative wage for commercial airline pilots than would be true in the absence of their unionization, with pilot quality not tightly specified in collective bargaining, would probably be accompanied by substantial increases in hiring standards for pilots by the airlines" (p. 46).

[12] This approach is of general applicability, see Anne Krueger.

[13] If pace of work times hours is the relevant "quantity" of labor services in utility and production functions, then an increased work pace would imply that the firm returns to the original (preunion) point on the factor-demand schedule and the workers return to the original utility level.

$$(17)\quad \ln W_i = \alpha_{11} + \beta_{11} Ed_i + \beta_{12} Exp_i + \beta_{13} Exp_i^2 + \beta_{14} Union_i + \Sigma\, \beta_{1j} Dem_{ji} + \Sigma\, \beta_{1k} WC_{ki} + \varepsilon_{11i}$$

$$(18)\quad Union_i = \alpha_{12} + \Sigma \gamma_{1k} WC_{ki} + \Sigma \gamma_{1j} Dem_{ji} + \Sigma \gamma_{1m} Z_{mi} + \varepsilon_{12i}$$

where W_i is the hourly wage rate of the ith individual; Ed_i is the years of education of the ith individual; Exp_i is the years of labor force experience of the ith individual; $Union_i$ is the union membership status of the ith individual; Dem_{ji} is the jth demographic characteristic of the ith individual; WC_{ki} is the kth working condition variable measured for the ith individual; and where Z_{mi} is a set of exogenous variables affecting membership.

If the WC_k variables are scaled so that higher scores measure less desirable working conditions (for example, less individually flexible hours, faster pace), then the compensating wage differential hypothesis leads to the expectation that $\beta_{1k} > 0$ and the interdependency explanation of unions leads to the prediction that $\gamma_{1k} > 0$. Several analyses (see Rosen; Stafford; Adrian Throop; Leonard Weiss) ignore (18), and in addition omit WC from (17). Estimates of the union effects from these studies (β_{14}) are generally higher than the 10–15 percent baseline estimate of Lewis.

Note that model (17)–(18) is recursive. Union status is an intervening variable between working conditions and wages, and the total effects of the working conditions on the wage can be allocated into *direct* effects that operate independently of union status (β_{1k}) and *indirect* effects that operate through union status ($\gamma_{1k} \cdot \beta_{14}$). Further, we can ignore union supply effects if it is assumed that union services are supplied at constant cost to all potential demanders. Equation (18), of course, is only one possible form of a membership equation. Some recent studies have used cross-sectional aggregate data (Orley Ashenfelter and Johnson) and data on individuals (Peter Schmidt and Robert Strauss) to test structural models which include an equation with union membership dependent upon wage rate as well as exogenous variables such as industrial concentration.[14] While we agree that it is important to incorporate the determinants of union membership into the model, we believe that union status is not a function of high wage rates per se, but rather is a response by workers to the character of work in certain jobs or occupations. Neither time-series nor cross-section evidence suggests a strong relationship between wages and union membership. Since 1950, real wages have risen steadily but the fraction of the work force that is unionized has fallen, and the highest paying occupations are not those with the highest rates of unionization.

Empirical studies of worker motivation in joining unions are consistent with the notion that working conditions are important.[15] Because of the heterogeneity across types of unions, there will certainly be some unions which are not formed to deal with issues in the organization of work, and our interdependencies hypothesis seems most appropriate for industrial unions. Although our sample size limits us from attempting disaggregations by type of union (as noted in Section III), Howard Foster has presented evidence that construction, which represents a major share of craft union members, has remained highly unionized in the commercial sector while unionism has waned in the residential sector. This is because of higher organizational costs for many small, geographically dispersed sites in the suburbs, which now represent the dominant part of residential construction. Commercial construction is likely to involve a more complex organization in the work setting.

---

[14] In addition, Ashenfelter and Johnson have a third equation which has labor quality as its dependent variable.

[15] See Joel Seidman, Jack London, and Bernard Karsh. They find that none of their sample of 113 members of a large midwestern United Steelworkers local reported joining to get higher wages. Most joined after unpleasant experiences at plants.

In terms of our theoretical discussion, a more reasonable objection to the model specified in equations (17)–(18) is that working conditions are not exogenous, but instead are likely to be affected by unions via some demand-revealing mechanism. Union contracts often have specific provisions concerning working conditions. In addition, as argued earlier, unions may increase the wage rate above competitive levels and induce employers to alter such working conditions as work pace. A system of equations which takes into account the simultaneity between working conditions and unions is as follows:

$$(19)\quad \ln W_i = \alpha_{21} + \beta_{21} Ed_i + \beta_{22} Exp_i$$
$$= + \beta_{23} Exp_i^2 + \beta_{24} Union_i + \Sigma\, \beta_{2j} Dem_{ji}$$
$$+ \Sigma\, \beta_{2k} WC_{ki} + \varepsilon_{21i}$$

$$(20)\quad Union_i = \alpha_{22} + \Sigma \gamma_{2k} WC_{ki} + \Sigma \gamma_{2j} Dem_{ji}$$
$$+ \Sigma \gamma_{2j} Z_{mi} + \varepsilon_{22i}$$

$$(21)\quad WC_{ki} = \Psi_{2k} + \delta_{2k}\, Union_i$$
$$+ \Sigma \phi_{2k} X_{ni} + \varepsilon_{2ki}$$

where the X_n are identifying variables in the working condition equations.

While it still is expected that γ_{2k}, β_{2k}, and $\beta_{24} > 0$, the direction of the effects of unions on working conditions (δ_{2k}) is indeterminate and will depend upon the union's willingness to trade off wages for favorable working conditions under the interdependencies hypothesis and firm's response to union (wage) demands in the partial control hypothesis.

Although the model specified in equations (19)–(21) is more complete than the model in (17)–(18), its data demands are extraordinary.[16] Reliable estimates of the effects of unions on working conditions could be obtained from longitudinal data on working conditions by industry both before and after unionization. Lacking such data and being unsuccessful in one attempt to estimate (19)–(21) with cross-sectional data on individuals, we present in the next section two sets of empirical results. In the first, we use cross-sectional data to estimate the recursive system of equations specified in (17)–(18). In the second, we use longitudinal data on individuals to compare changes in wages and changes in a limited set of working condition variables for blue-collar workers who joined unions, quit unions, or remained in either union or nonunion work. We find strong evidence that those joining unions increase their wages at the expense of desirable working conditions.

III. Empirical Results

A. *Data and Variable Definitions*

Our major data sources are the 1975–76 *Time Use Survey*, conducted by the Survey Research Center between October 1975 and October 1976, and the *Panel Study of Income Dynamics: Procedures and Tape Codes*, also conducted by the Survey Research Center.[17] In the former, four interviews were taken with a representative sample of adults (and their spouses) at four-month intervals. Most of the *Time Use Survey* data used here come from the second re-interview, which took place during May and June 1976. Of the 1,007 third-wave respondents, we selected those who had regular work schedules, gave valid responses to the wage and work hours questions, and reported working in blue-collar occupations. The resulting sample consists of 149 observations. Because this sample size is small relative to most other micro-data sets, we spent a considerable amount of time checking inconsistent responses and other errors in the data.[18]

The variables used in estimating equations (16) and (17) fall into four groups: wage rates, labor quality, union status, and working conditions. A brief description of specialized variables follows.

---

[16] Further, some authors have postulated that a wage variable belongs in (21). This was suggested to us by Henry Farber.

[17] We also use the 1972–73 *Quality of Employment Survey*.

[18] The major reason for the small sample size is because the *Time Use Survey* was designed to measure adult time use regardless of labor market status.

Wage Rate. Our data are unique in that they allow quantification not only of nominal wage rate but also wage rates that are adjusted for time spent in work breaks—formal and informal—as well as time spent in on-the-job training. The "nominal wage rate" is defined as weekly earnings divided by nominal work hours.[19] After subtracting time spent in formal and informal work breaks[20] from nominal work hours, we calculated "wage-break time" by dividing weekly earnings by this adjusted measure of hours. A final adjustment to the wage rate was made possible by information on hours spent in on-the-job training. Two types of training are distinguished—training occurring jointly with usual work activities and training taken separately from usual work but during the work period.[21] Our measure of total weekly training time was the sum of the hours of separate training and one-third of the hours reported in joint training. When weekly earnings were divided by an hours measure which netted out both break time and on-the-job training, we obtained our third measure of wage rate—"wage-break and training time."

Working Conditions. Machine = one if the respondent reports almost all or most of the time on the job is involved with using machines, and zero otherwise. *Not Free* = zero if the respondent reports that it would be "easy to get a couple of hours off from work" to do shopping or personal business and one otherwise. *ln Work Effort* = the natural logarithm of a scale of work effort formed from responses to questions asking for a rating, on a zero to ten scale, of the amount of energy and effort put into "a typical hour of time at work" and "a typical hour of time spent watching TV." The variable is defined by[22]

$$\frac{Work\ \# - TV\ \#}{10 - TV\ \#}$$

The upper bound for the scale is 1.0, and for those few cases in which the above expression equaled zero, the value was set to .01. The scale was multiplied by 100, and the resulting mean of the natural logarithm is 4.0.

Union Status. Union status is measured as a dichotomous variable, equaling one if the worker reported being a union member and zero otherwise.

B. *Estimation*

The estimated coefficients of a basic earnings equation (17) for each of the three wage rate measures are given in Table 1. The working condition variables are excluded from the regression results reported in the first three columns and are included in the

---

[19] Nominal hours were ascertained from the following question: "About how many hours do you work on your main job in an *average* week including any paid or unpaid overtime?"

[20] The question for formal breaks was, "About how much time do you usually spend on regular coffee breaks or scheduled rest breaks," and for informal breaks, "Thinking about the *rest* of your time *at work*, about how much (additional) time do you usually spend on things not related to the work that you do—like talking to friends, doing personal business, or just relaxing?" To the sum of these two responses we added time spent in lunch breaks which exceeded 60 minutes.

[21] The question sequence was as follows: "Do you feel you are learning skills on your job that could lead to a better job or to a promotion?" (If Yes) "Sometimes people learn these skills as part of their regular work, while others use time at work to learn skills that are *not* part of their regular job. About how many hours *per week* do you usually spend learning new things as *part of* your regular work?" "And how many hours *per week* do you spend learning new things that are *not* part of your regular work?"

[22] The scale may require some additional explanation. A simple 10-point scale assessment of work effort can be faulted because of respondent-to-respondent differences in use of the scale. Some tend to cluster their responses around the midpoint of the scale *regardless* of the activity, while others place their responses at the extremes of the scale. Since television watching is a nearly universal activity which requires nearly uniform (lack of) effort, the effort rating of television was used to anchor the zero point of our work effort variable. In the small number of instances that the work effort rating was less than or equal to television effort, the ratio was set equal to 0. When both activities were scored "10" (also very rare), the ratio was set equal to unity. In our empirical work, we experimented with several alternative measures of work effort and found the natural logarithm of the ratio to be superior to the alternatives, especially the simple 10-point assessment of work effort.

TABLE 1—ESTIMATED REGRESSION COEFFICIENTS FROM VARIOUS WAGE EQUATIONS, FOR BLUE-COLLAR WORKERS

	Dependent Variable					
Independent Variable	*ln* Nominal Wage Rate	*ln* Wage-Break Time	*ln* Wage-Break and Training Time	*ln* Nominal Wage Rate	*ln* Wage-Break Time	*ln* Wage-Break and Training Time
Education years	.040[b]	.038[a]	.051[b]	.041[a]	.040[a]	.054[b]
	(.018)	(.019)	(.020)	(.017)	(.018)	(.020)
Experience (years)	.029[a]	.027[a]	.015	.021	.019	.009
	(.012)	(.013)	(.014)	(.012)	(.013)	(.014)
Experience^2	−.00053	−.00047	−.00027	−.00038	−.00034	−.00016
	(.00028)	(.00030)	(.00032)	(.00028)	(.00030)	(.00032)
Female	−.386[b]	−.442[b]	−.486[b]	−.408[b]	−.456[a]	−.481[b]
	(.090)	(.095)	(.101)	(.088)	(.093)	(.100)
Nonwhite	−.088	−.085	−.029	−.037	−.034	−.010
	(.141)	(.149)	(.159)	(.138)	(.146)	(.158)
City Size (if 50,000 or more)	.133	.137	.152	.130	.139	.160
	(.084)	(.089)	(.094)	(.083)	(.088)	(.095)
Northeast	.204	.180	.281	.186	.149	.247
	(.138)	(.145)	(.154)	(.134)	(.143)	(.154)
North Central	.231[a]	.219	.236[a]	.200	.169	.182
	(.106)	(.111)	(.118)	(.105)	(.111)	(.120)
West	−.144	−.169	−.194	−.162	−.204	−.233
	(.122)	(.128)	(.137)	(.120)	(.127)	(.137)
Machine	–	–	–	.214[b]	.229[b]	.199[a]
				(.084)	(.089)	(.096)
Not Free	–	–	–	.067	.122	.132
				(.090)	(.096)	(.103)
ln Work Effort	–	–	–	.078[a]	.045	.011
				(.032)	(.034)	(.037)
Constant	.819	1.009	1.031	.459	.744	.894
R^2/SEE	.313/.487	.309/.512	.302/.544	.369/.471	.354/.501	.330/.539
N	149	149	149	149	149	149

Source: 1975–76 Time Use Study.

Note: Standard errors are in parentheses.

[a] Significant at the .05 level.

[b] Significant at the .01 level.

regressions of the final three columns. Overall, the results show some support for the hypothesis that working condition variables lead to compensating wage differentials. All of the estimated coefficients for these three variables are positive, although only for the machine variables are the coefficients consistently significant at conventional levels. The work effort variable is significant when related to the nominal wage rate, although it becomes insignificant when hours spent in breaks and training are subtracted from work hours. The coefficient of the not free variable exceeds its standard error in both cases when the wage rate is adjusted for nonwork time.[23] There is also some tendency for the coefficient on education to increase with the adjustments indicating that those with high education receive more training and more work breaks in these blue-collar jobs.[24]

When the union status variable is added to these equations, the results (presented in the first six columns of Table 2) show a decline in estimated union effects in the presence of adjustments to the wage rate

---

[23] A positive relation between bad working conditions and wages is *not* typical for cross-sectional analysis. Negative coefficients are often interpreted as evidence of omitted skill variables combined with pleasant working conditions as normal goods.

[24] We were unable to reject the hypothesis of structural equality when regressions were run separately for men and women. However, it can be noted that the earnings disadvantage of women grows with more refined wage measures. This is because women spend less time in work breaks and on-the-job training.

TABLE 2—ESTIMATED REGRESSION COEFFICIENTS FOR UNION STATUS AND WORKING CONDITION VARIABLES FOR VARIOUS WAGE EQUATIONS, FOR BLUE-COLLAR WORKERS

Independent Variable	*ln* Nominal Wage Rate	*ln* Wage-Break Time	*ln* Wage-Break and Training Time	*ln* Nominal Wage Rate	*ln* Wage-Break Time	*ln* Wage-Break and Training Time	*2SLS* (replication of A-J) *ln* Wage-Break and Training Time
Union Status	.272[b]	.268[b]	.235[a]	.187	.185	.174	.213
	(.092)	(.097)	(.104)	(.097)	(.104)	(.112)	(.510)
Machine	–	–	–	.169[a]	.184[a]	.157	.143
				(.086)	(.092)	(0.99)	(.144)
Not Free	–	–	–	.031	.085	.098	–
				(.092)	(.097)	(.105)	
ln Work Effort	–	–	–	.066[a]	.033	.000	*EXOG*
				(.032)	(.035)	(.037)	
Constant	.793	.983	1.008	.511	.796	.943	.986
R^2/SEE	.353/.474	.345/.500	.327/.536	.386/.467	.369/.500	.342/.536	.338/.534
N	149	149	149	149	149	149	149

Source: See Table 1.

Note: Other variables included in all but the two-stage least squares (*2SLS*) regression equations: Education, Experience, Experience^2, Female, Nonwhite, City Size, Northeast, North Central, West. In the *2SLS* equation, Education, Experience, Experience^2, Female, City Size, the three-region dummy variables, and Nonwhite were included exogenous variables. *Ln* Work Effort was the excluded exogenous variable. *EXOG*=exogenous variable excluded from equation. Standard errors in parentheses.

[a], [b]See Table 1.

and with the addition of the working condition variables. The decline in the estimated union-nonunion differential from a 31.3 percent[25] differential in nominal wages to a 26.5 percent difference when break and training time adjustments are made comes about because union members are found to have slightly less break time and substantially less time devoted to on-the-job training. For our sample of blue-collar workers, union members report fewer hours of both joint and separate on-the-job training. The reports of weekly training hours that are a part of regular work average 3.8 hours for union members and 4.6 hours for nonmembers. The difference for training that is *not* a part of regular work is proportionately larger—0.4 vs. 1.5 hours per week.[26] We interpret these differences as support for our view that organization of work can in part substitute for specific individual skills.

In addition to adjustments to relate the hourly wage more closely to hours of actual work time (cols. 1–3 of Table 2), we introduced the working condition variables (cols. 4–6). When the working condition variables are added, the estimated union effect becomes statistically insignificant although its point estimate (18.9 percent) is still larger than the 10–15 percent range of Lewis' estimates.[27] In the final column of Table 2 are two-stage estimates of a simultaneous equation model analogous to that of Ashenfelter-Johnson (A-J). With this formulation, the estimated union-nonunion dif-

---

[25] This is calculated $e^{\hat{\beta}}-1$, where $\hat{\beta}$ is the estimated coefficient on the "union status" variable given in Table 2.

[26] If we had a larger sample we would expect to observe differences among different occupations. Union craftsmen in nonindustrial settings would likely engage in more on-the-job training than their nonunion counterparts.

[27] Although a sex-union status variable was not statistically significant when added to the equations reported in Table 2, estimated union effects were considerably lower for men than women. The six coefficients for the union status variable for men only were (with standard errors in parentheses): .186 (.105), .177 (.110), .130 (.120), .122 (.112), .101 (.119), and .062 (.129). Thus, the point estimate of the union effect on the adjusted wage measure *without* the working condition variables is 13.9 percent and the estimate based on the regression *with* working condition variables is 6.4 percent.

TABLE 3—ESTIMATED REGRESSION COEFFICIENTS FROM VARIOUS UNION MEMBERSHIP EQUATIONS, FOR BLUE-COLLAR WORKERS
(Dependent Variable = Union Status)

Independent Variable	*OLS* Linear Probability	Logit	Replication of A-J Membership Equation	
			OLS	*2SLS*
Education (years)	−.002	−.001	–	*EXOG*
	(.015)	(.058)		
Experience (years)	.025^{a}	.082^{a}	–	*EXOG*
	(.011)	(.038)		
Experience^{2}	−.00064^{b}	−.0021^{a}	–	*EXOG*
	(.00025)	(.0009)		
Female	−.212^{b}	−.743^{b}	–	*EXOG*
	(.076)	(.288)		
Nonwhite	−.028	−.077	.045	.063
	(.120)	(.500)	(.121)	(.127)
City Size (if 50,000 or more)	.017	.096	−.005	−.042
	(.072)	(.253)	(.074)	(.080)
Northeast	.111	.361	.049	−.030
	(.117)	(.387)	(.120)	(.132)
North Central	.192^{a}	.589	.156	.108
	(.092)	(.307)	(.094)	(.102)
West	.037	.073	.039	.072
	(.104)	(.386)	(.105)	(.111)
Machine	.241^{b}	.794^{b}	.256^{b}	.204^{a}
	(.073)	(.257)	(.073)	(.082)
Not Free	.197^{a}	.631^{a}	.164^{a}	.148
	(.079)	(.269)	(.080)	(.084)
ln Work Effort	.065^{a}	.227	.061^{a}	.061^{a}
	(.028)	(.122)	(.029)	(.030)
ln Wage adjusted for break and training time	–	–	.161^{b}	.375^{b}
			(.059)	(.133)
Constant	−.278		−.456	−.759
R^{2}/*SEE*	.286/.411		.235/.421	.202/.440
n	149	149	149	149

Source: See Table 1.

Note: *EXOG* = Exogenous variable excluded from equation. Standard errors in parentheses.

[a], [b] See Table 1.

ferential is 21.7 percent and statistically insignificant, although we found that the estimated union coefficient is sensitive to the exogenous variables omitted from the equation.

Estimates of union membership equations are presented in Table 3. In the first and second column of numbers, estimates of equation (18) (in which union membership is taken to be a function of working conditions but not wages) are presented. The functional forms assumed for these two membership equations are linear probability and logit, respectively. Regardless of functional form, the three working condition variables have strong, positive effects on the likelihood of belonging to a union. In the linear probability equation, working with machines as a proxy for capital intensity increases the probability of being a union member by about 25 percent, being constrained with free time increases it by close to 20 percent, and a 1 percent increase in the work effort variable is associated with a 6.5 percent increase in the likelihood of belonging to a union.

Several of the demographic variables are also significant in the membership equation. There is a significant parabolic membership-experience relationship, peaking at

TABLE 4—CHANGE IN WAGE RATES AND WORKING CONDITIONS, 1968–71, BY CHANGE IN UNION STATUS
(All Blue-Collar Male Workers)

Union Status Change 1968 and 1971	Absolute Change in Wage Rate	Percentage Change in Wage Rate	Change in Choice in Work	Change in Freedom to Increase Work Hours	Weighted Percent of Observations	Number of Observations
Joined Union	\$1.28	54.8	−.07	−.28	8.0	114
Quit Union	0.12	7.0	.06	−.08	5.6	78
Remained in Union	1.14	32.6	.01	−.07	39.1	434
Remained nonunion	0.73	39.9	.02	−.08	47.3	730
Mean	\$0.90	36.4	.01	−.09	100.0	1,356
Eta^{2d}	.027^{b}	.010^{b}	.005^{c}	.007^{a}		

Source: Panel Study of Income Dynamics.
[a], [b]See Table 1.
[c]Significant at .10 level.
[d]Fraction of variance in dependent variable explained by Union Status Change categories.

nineteen years of work experience. Females have significantly lower rates of union membership. Workers residing in the North Central region are significantly more likely to be union members than otherwise similar workers in the southern region. The final two columns of Table 3 are an attempt to replicate the membership equation of Ashenfelter and Johnson. The adjusted wage rate is included and the equations are estimated with ordinary least squares and two-stage squares, respectively. The results show that *ln* wage rate has a significant effect on union membership. The coefficients on the work pace variables drop somewhat, although, with the exception of the not free variable in the two-stage equation, they are still statistically significant at the 5 percent level.

If the relationship among work pace, union status, and wage rate is recursive, as assumed in equations (17) and (18), the total effects of the working condition variables can be allocated into *direct* effects operating independently of union status and indirect effects which exist because the working conditions are hypothesized to cause unions, and unions in turn cause higher earnings. The total effects are shown in the final column of Table 1. Of greatest interest is the machine variable, which has total effects that are statistically significant. As shown in the sixth column of Table 2, about four-fifths of this total effect is direct, operating independently of union status, leaving about one-fifth of the total effect to operate through union status.

The cross-sectional *Time Use Survey* data do indicate the presence of compensating wage differentials for union members. Because cross-sectional results are particularly susceptible to error due to omitted variables —in this case, omitted measures of worker quality—we sought confirmation of the compensating wage differential result with longitudinal data. The source of our longitudinal data is the first five waves of the *Panel Study*. The sample was restricted to male household heads who worked at least 250 hours per year between 1967 and 1971 and who had blue-collar jobs in the spring of 1968. The *Panel Study* data provide information on two working condition variables: choice in work and freedom to increase work hours.[28] Table 4 shows changes in union status. Those joining unions enjoyed the largest increase in hourly earnings, but also experienced the largest declines in the working condition variables. Those quitting unions had the smallest wage increases, but also had the highest average gain in reported choice in work. Thus, the hypothesis that union members receive compensating wage differentials is given strong support in these longitudinal data as well.

---

[28] The choice in work variable was constructed from the question, "How much choice do you have about the way you do the work?" Responses were coded in quintiles as percent of choice and this variable was set equal to the midpoint of the quintiles, taking values of .1, .3, .5, .7, and .9. The freedom to increase work hours variable is dichotomous and is constructed from the question, "Was there more work available on your job so that you could have worked more if you had wanted to?" Positive responses were scored 1 and all other responses were scored zero.

IV. Discussion

We believe the evidence presented from the *Time Use Survey* supports our view that unions are more prevalent in work settings which require interdependent worker behavior as part of the production process. Though we have no direct evidence, we believe that consumption aspects of work are also shared by most workers on jobs which involve greater interdependencies. Specific variables which were designed to measure work pace and extent of scheduling were found to be important predictors of union membership for blue-collar workers. Working with machines was also an important predictor of union membership. Working with machines is likely to imply a production process which has more capital per worker, and hence there will be greater incentives to keep the machinery operating on a continuous basis. This will tend to reduce the possibility of unscheduled breaks and will increase the possibility of work shifts or longer workdays.[29] If many workers are subject to a single work group or plant or company-wide policy on work pace and work scheduling to accomplish the fuller use of the capital stock, we expect and observe a higher probability of some less individualistic mechanism, including unionism, for reaching agreement on such issues.

Although some unions have occupational licensure as a major function, in our view unions can represent a mechanism for reaching collective agreements about the organization of work. Collective agreement is reached because search in the labor market to match up workers and firms who "agree" on several major aspects of the organization of work is likely to be more costly than organizing a union and paying a high enough wage to compensate for the inherent restrictions the work setting will place on at least some individuals.[30] Transaction and search costs will thereby generate rents for some workers though unions can reduce the level of these rents by better identifying preferences of the "median worker." As noted by Richard Freeman, up to a point, "voice" can be more efficient for resource allocation than "exit" even from the perspective of the employer.[31]

One criticism of our approach is similar to that which we raised concerning the hypothesis that unionism is a function of wages per se. Namely, if membership depends on these certain variables describing the nature of work, why don't we observe a universal relation between certain working conditions and unionism? For instance, why aren't white-collar workers organized into unions when the organization of work entails substantial interdependencies in work effort? One answer is traditional; white-collar workers are more likely to aspire to management positions and for this reason will not be interested in participation in union activities. Our answer is that those white-collar workers whose jobs have working conditions which we found to be important predictors of blue-collar unionism are also the most likely to be organized. In a regression of union membership for the 218 white-collar workers the equivalent of the logit equation in Table 3 is

---

[29] All three working condition variables are positively correlated.

[30] Company unions could perform many of these functions. In Korea, public policy severely restricts the activity of automous labor unions and company unions have grown rapidly in the manufacturing sector. As noted by Y. K. Park: "In recent years company based unions have become increasingly common in some of the rapidly expanding sectors of manufacturing. This trend is independent of government encouragement or discouragement but is consistent with firm objectives"(p. 57).

[31] John Dunlop has argued that "...the institutionalized form of collective action may introduce new preferences in the same way the household modifies individual preferences" (p. 30). Interaction patterns of a given form of collective action could alter preferences and in addition various forms of collective action (for example, voting rules) can obviously affect the choice of the group even if preferences remain stable.

$$ln \frac{P(Union=1|X_i)}{P(Union=0|X_i)} = \underset{(.256)}{.135}\ Machine + \underset{(.228)}{.541}\ Not\ Free + \underset{(.274)}{.437}\ ln\ Work\ Effort + \ldots$$

where X_i represents the vector of independent variables.

The coefficients for not free and *ln* work effort are of the same general magnitude of the corresponding coefficients in the blue-collar equation (.631) and .227, respectively) though the standard errors are proportionately higher. What is notable is that these two variables are the only ones having any significant relation to union membership for white-collar workers. The insignificance of the machine variable for white-collar workers is not surprising since the machine is often a simple piece of office equipment (notably typewriters). As such, working with machines is less likely to imply a high ratio of capital per worker for white-collar workers. Unlike the case of blue-collar workers, however, one cannot observe compensating wage differentials for these nonpecuniary features of work. We suspect that the omitted variable problem stated in Section III is dominating the empirical results, and that white-collar workers with high potential wages choose to buy better working conditions, particularly in light of progressive marginal tax rates.

Because the central hypothesis of this paper is that organization of work is related to unionism we attempted to replicate our blue-collar union membership equation on data from the *Quality of Employment Survey*.[32] These data contain measures of overtime-hours flexibility and the extent to which the respondent characterized his job as allowing freedom as to how to do work, requiring hard work, physical effort, allowing decision making or being repetitive. When these variables and demographic controls were related to a dichotomous measure of union status in a linear probability regression for 570 nonfarm blue-collar workers with regular work schedules, several of the working condition variables were statistically significant, and all but one of these had the expected sign. Among the significant variables were two dichotomous overtime-hours flexibility variables, measuring whether overtime work was set by a supervisor, and if so, whether it could be refused without penalty. Those who would be penalized for refusing overtime work were 17 percent more likely to belong to a union than those who could set their own overtime hours, while those who wouldn't be penalized but did not have control over their overtime hours were 13 percent more likely to be union members than those with freedom to determine their own overtime hours. Respondents who reported having blue-collar jobs which allow freedom "as to how you do your work" were significantly *less* likely to be labor union members. The work effort variables produced mixed results on union membership. Respondents reporting jobs which require "a lot of physical effort" were *more* likely to be members (significant at the 10 percent level) but those reporting jobs which require "you to work very hard" were significantly *less* likely to be union members. This latter result is the only one that runs contrary to our expectations, and we have no ready explanation for it. We do note that this work effort variable, however, is considerably less refined than the one constructed from the *Time Use Survey* data. On the whole, our hypothesis that unions are a function of working conditions holds up quite well on this larger data set.

Since blue-collar unionism is more prevalent in jobs with certain working conditions and since these working conditions require higher wages, the apparent effect of unions on wages is reduced once allowance is made for these working conditions. From our ordinary least squares model (Table 2) the relative wage advantage falls from 31.3 to 18.9 percent for all blue-collar workers and falls from 20.4 to 6.4 percent for male blue-collar workers.

---

[32] Interviews were taken with a nationally representative sample of working adults.

REFERENCES

O. Ashenfelter and G. Johnson, "Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries," *Int. Econ. Rev.*, Oct. 1972, *13*, 488–508.

H. R. Bowen, "The Interpretation of Voting in the Allocation of Economic Resources," *Quart. J. Econ.*, Nov. 1943, *58*, 32–42.

F. Brechling, "The Relationship Between Output and Employment in British Manufacturing Industries," *Rev. Econ. Stud.*, July 1965, *32*, 187–216.

E. H. Clarke, "Multipart Pricing of Public Goods," *Publ. Choice*, Fall 1971, *11*, 17–33.

R. Coase, "The Nature of the Firm," *Economica*, Nov. 1937, *4*, 386–405.

A. V. Deardorff and F. P. Stafford, "Compensation of Cooperating Factors," *Econometrica*, July 1976, *44*, 671–84.

Howard Foster, *Manpower in Homebuilding: A Preliminary Analysis*, Philadelphia 1974.

R. B. Freeman, "Individual Mobility and Union Voice in the Labor Market," *Amer. Econ. Rev. Proc.*, May 1976, *66*, 361–68.

T. Groves and J. Ledyard, "Some Limitations of the Demand Revealing Process," *Publ. Choice*, Spring 1977, suppl., *28*, 107–124.

G. E. Johnson, "Economic Analysis of Trade Unionism," *Amer. Econ. Rev. Proc.*, May 1974, *64*, 23–28.

A. O. Krueger, "The Political Economy of the Rent Seeking Society," *Amer. Econ. Rev.*, June 1974, *64*, 291–303.

H. M. Levinson, "Unionism, Concentration and Wage Changes: Toward a Unified Theory," *Ind. Labor Rel. Rev.*, Jan. 1967, *20*, 198–205.

H. Gregg Lewis, *Unionism and Relative Wages in the United States*, Chicago 1963.

Y. K. Park, *Labor and Industrial Relations in Korea: System and Practice*, Seoul 1974.

S. Rosen, "Unionism and Occupational Wage Structure in the United States," *Int. Econ. Rev.*, June 1970, *11*, 269–86.

______, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *J. Polit. Econ.*, Jan./Feb. 1974, *82*, 34–55.

P. A. Samuelson, "A Diagrammatic Exposition of a Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1955, *37*, 250–56.

P. Schmidt and R. Strauss, "The Effect of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach," *Int. Econ. Rev.*, Feb. 1976, *17*, 204–12.

S. J. Seidman, J. London, and B. Karsh, "Why Workers Join Unions," *Annals Amer. Acad. Polit. Soc. Sci.*, Mar. 1951, *274*, 75–84.

F. P. Stafford, "Concentration and Labor Earnings: Comment," *Amer. Econ. Rev.*, Mar. 1968, *58*, 174–80.

George J. Stigler, *The Theory of Price*, 3rd ed., New York 1966.

C. M. Tiebout, "A Pure Theory of Local Expenditure," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

A. W. Throop, "The Union-Nonunion Wage Differential and Cost Push Inflation," *Amer. Econ. Rev.*, Mar. 1968, *58*, 79–99.

L. Weiss, "Concentration and Labor Earnings," *Amer. Econ. Rev.*, Mar. 1966, *56*, 96–117.

Survey Research Center, *A Panel Study of Income Dynamics: Procedures and Tape Codes*, Vols. I and II, Inst. Soc. Res., Ann Arbor 1974.

______, *1972–73 Quality of Employment Survey*, SSA Study Number 3510, SCR Study Number 462330, Inst. Soc. Res. Soc. Sci. Archive, Ann Arbor 1975.

______, *Time Use in Economic and Social Accounts*, Inst. Soc. Res., Ann Arbor 1976.

The Marginal Utility of Income Does Not Increase: Borrowing, Lending, and Friedman-Savage Gambles

By MARTIN J. BAILEY, MANCUR OLSON, AND PAUL WONNACOTT*

There has been a great deal of discussion about whether the marginal utility of income rises with income. Most notably, Milton Friedman and Leonard J. Savage argued in their classic paper that the willingness to gamble implies that the marginal utility of income is rising over a range. The discussions of rising marginal utility and of Friedman-Savage gambles have proceeded independently of the literature on time preference, although the issues are in fact related logically. Drawing on this logical relationship we shall show 1) that at least when intertemporal utility is separable, the stable levels of consumption that are usually observed imply that the marginal utility of income decreases as income rises. 2) Even if the marginal utility of income does increase, Friedman-Savage gambles normally will not maximize utility; saving and dissaving can attain the levels of consumption which generate the most utility per dollar of income at a lower cost than gambles unless imperfections in the capital market are severe. 3) Even when the utility function is not temporally separable, repeated gambling cannot be a rational way of dealing with a rising marginal utility of income. We therefore conclude that observed gambling is seldom if ever explained by the logic set out in Friedman and Savage's seminal paper. In view particularly of the stability of consumption levels and the lack of Friedman-Savage gambles, we conclude that marginal utility of income does not rise with income.

I. A Conceptual Framework

We begin with the assumption of separable utility.[1] In this we follow the lead of Paul Samuelson, Robert Strotz, Gerard Debreu, Tjalling Koopmans, and Jack Hirshleifer. With separable utility we can represent current utility in time period t by a function $v(C_t)$. Consumption is time subscripted, not the function itself;[2] it depends only on the quantity C_t, regardless of t. Subject to a wealth budget constraint, the consumer maximizes the discounted value of present and future single period utilities

$$U=\sum_{t=0}^{T}\frac{v(C_t)}{(1+\eta)^t} \tag{1}$$

where the consumer's time horizon for planning is T, which need not be finite, and where η is the rate of pure time preference. We assume now that the first derivative $v'(C_t)$ is positive and that η is a positive constant. Olson and Bailey have demonstrated in another paper (1981) that η is definitely positive, and thus for various reasons there is a preference for present over future consumption that would be obvious if the interest rate were zero.

We write the maximization problem using the customary Langrangian

$$U^*=U+\lambda\left[\sum_{t=0}^{T}\frac{Y_t-C_t}{(1+r)^t}\right] \tag{2}$$

---

*Professors of economics, University of Maryland. We are grateful to Roger Betancourt, Christopher Clague, Samuel B. Chase, Jr., John Harsanyi, Howard Margolis, and Paul Meyer for helpful comments at various stages, but are alone responsible for any shortcomings. Part of the work on this article was supported by grants from the National Science Foundation and Resources for the Future.

[1] Bailey and Olson argue (1977, 1978) that the assumption of temporal separability is not only a fruitful simplifying assumption but is closer to being descriptively accurate than some economists have assumed.

[2] Changes in health, age, family status, or other conditions that might alter the value of consumption in particular periods can be accommodated by adding state variables. Since such variables would not affect the conclusions here, we do not include them now.

where income in period t is Y_t, and where r is the rate of interest. We ignore bequests, treating them in effect as extensions of the planning horizon T. Suppose income is a constant over time, Y; then we can write utility as a function of income

$$U^* = U + \lambda\left[\sum_{t=0}^{T} \frac{Y - C_t}{(1+r)^t}\right] \tag{3}$$

When the consumer maximizes utility, he uses up his lifetime income in lifetime consumption, so that the term in brackets is always zero when U^* is maximized. Let $U^{**} = \max U^*$. Therefore it works out that

$$\frac{dU^{**}}{dY} = \sum_{t=0}^{T} \frac{\partial U}{\partial C_t}\frac{dC_t}{dY} = \sum_{t=0}^{T} \frac{v'(C_t)}{(1+\eta)^t}\frac{dC_t}{dY} \tag{4}$$

Now consider the special case in which the consumer cannot borrow, lend, or gamble. In this case, he must always consume exactly his income:

$$C_t \equiv Y \tag{5}$$

so that $dC_t/dY = 1$. Then equation (4) becomes

$$\frac{dU^{**}}{dY} = \sum_{t=0}^{T} \frac{v'(Y)}{(1+\eta)^t} \tag{6}$$

and we have

$$\frac{d^2U^{**}}{dY^2} = \sum_{t=0}^{T} \frac{v''(Y)}{(1+\eta)^t} = v'' \sum_{t=0}^{T} \frac{1}{(1+\eta)^t} \tag{7}$$

The expression (7) is positive or negative according as $v''(Y)$ is positive or negative; the "marginal utility of income" is increasing (decreasing) if and only if the marginal utility of consumption is increasing (decreasing).

Note, however, that the marginal utility of income (or of wealth) is an imprecise concept because it does not take account of the constraints under which the consumer maximizes utility, for example, that in (5) or that in brackets in (2). If the consumer can borrow or lend at a given interest rate he will often get more utility out of a fixed amount of lifetime consumption than if he cannot. This is particularly the case for any consumer for whom the marginal utility of consumption is rising over a range, since no consumer who can borrow or lend at the going interest rate (or engage in fair gambles) will ever choose any consumption level within that range; he will attain a higher level of U^{**} at a combination of low and high consumption in different periods such that the marginal utility of consumption is nonincreasing at both levels of consumption. Moreover, when income increases for an individual whose marginal utility is at that point rising, he increases the fraction of his time that he spends at the higher level of consumption, keeping the two levels of consumption unchanged. Over the range of income between the two levels of consumption, the marginal utility of income or wealth is constant. (We develop these points in later sections.)

Thus to avoid ambiguity it is preferable to focus directly on the utility of the consumption that income or wealth make possible. This calls attention to the need to specify the constraints under which utility is maximized and also helps to show why the Friedman-Savage view of gambling cannot be rescued by interpreting it in terms of a utility function using permanent income or wealth as its argument.

II. The Utility Function Revealed by Two-Period Choices

We will now show that we can infer the utility function $v(C_t)$ from observed two-period choices, where the function is unique except for the choice of origin and of a unit of measure, that is, is unique to within a linear transformation. Bailey and Olson have shown (1978) that when the household has equal amounts of consumption goods in the two periods the slope of an indifference curve through that point in two-period consumption space is $-(1+\eta)$; and for every other point in this space, the slope of an indifference curve passing through it is given by the following equation:

$$\text{(8)}\qquad \left[\frac{dC_1}{dC_0}\right]_{U=\text{const}} = \frac{v'(C_0)}{v'(C_1)}(1+\eta)$$

The theory of revealed preference tells us that an indifference map can be inferred from household choices between pairs of bundles of goods (C_0, C_1). That is, without knowing beforehand the function $v(\cdot)$ or the constant η, we can discover the indifference map from a suitably designed set of household choices. Then by measuring the slope of an indifference curve of such an indifference map as it crosses the 45° line, where $C_0 = C_1$, we can obtain the value of the time-preference constant η. Then, for every other consumption point in the map, we can use this constant and equation (8) to obtain the ratio

$$\text{(9)}\qquad \frac{v'(C_0)}{v'(C_1)} = -\frac{1}{1+\eta}\left|\frac{dC_1}{dC_0}\right|_{U=\text{const}}$$

For example, if we hold C_1 constant and vary C_0, obtaining from (9) the value of the ratio of marginal utilities of C_0 and C_1 for every value of C_0, we obtain to within a multiplicative constant the marginal utility $v'(C_0)$ for every value of C_0. Then by integrating $v'(C_0)/v(C_1)$ over a range of values of C_0 with C_1 constant, we obtain $v(\cdot)$ to within a multiplicative constant and a constant of integration. Consider the function

$$\text{(10)}\qquad V(\cdot) = a + bv(\cdot)$$

where a and b are arbitrary constants. Then the ratio

$$\text{(11)}\qquad \frac{V'(C_0)}{V'(C_1)} = \frac{bv'(C_0)}{bv'(C_1)} = \frac{v'(C_0)}{v'(C_1)}$$

so that when we integrate the ratio on the right of (11) with C_0 as the variable of integration, the result cannot be distinguished from the corresponding integral of the ratio on the left of (11). In other words, the form of $v(\cdot)$ obtained in this way is unique to within a linear transformation.

Thus, choices between distinct bundles of present and future consumption can be used to reveal the utility function $v(\cdot)$, which should be closely related to the function showing the utility of income revealed by choices involving gambles. The precise way in which they are related depends on assumptions about borrowing and lending opportunities and about opportunities to gamble. Experiments could be constructed so that the two functions would be identical to within a linear transformation, as we can see by considering the form and derivation of our equation (7). This finding sets the stage for our critique of the Friedman-Savage conclusions about gambling.

III. The Implications of Increasing Marginal Utility with Separability

In the analysis up to this point we have assumed that the marginal utility of consumption in any period decreases with consumption, that is, that $v(\cdot)$ is convex from above, so that $v''(\cdot) < 0$. Suppose instead that $v(\cdot)$ is concave, so that $v''(\cdot) > 0$ over a certain range of consumption, as shown in Figure 1. Then, as a consequence, the indifference curve is concave toward the origin over a range, as illustrated in Figure 2. Intuitively, the reason is most easily seen in the special case where $\eta = r = 0$. With rising marginal utility, uneven consumption (C_0, C_1) gives average utility of $(Q+R)/2 = S$ over the two-year period; that is, more

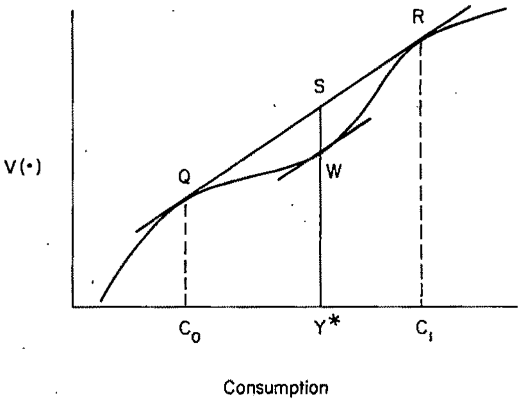


FIGURE 1

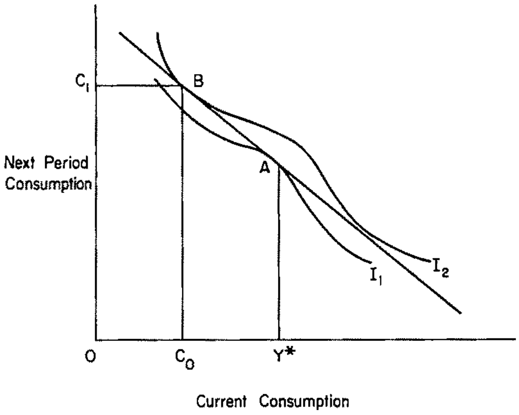


FIGURE 2

than the average of W that can be achieved by consuming the same amount each year. Thus, in Figure 2, B is preferred to the stable consumption at A.[3]

Rising marginal utility of consumption with separability thus implies highly unstable consumption levels, with households switching between higher and lower levels of consumption. But this is inconsistent with empirical evidence. Thus, if utility is separable, we may dismiss rising marginal utility of "income."

IV. Borrowing, Lending, Gambling, and "Risk Preference"

We shall now show that, even if the marginal utility of consumption is increasing, Friedman-Savage gambles can occur only in special circumstances because borrowing and lending will be preferable. Suppose that a household faces a budget constraint of

$$C_0 - Y^* = \frac{Y^* - C_1}{1+r} \tag{12}$$

where Y^* is the household's present permanent income, and where r is a fixed rate of interest at which the household can either borrow or lend from the current period until the next. Treating consumption after period 1 as given and fixed with a present utility of V, the household maximizes

$$U' = v(C_0) + \frac{v(C_1)}{1+\eta} + V \tag{13}$$

subject to the constraint (12) Solving this constrained maximization problem gives

$$\frac{v'(C_1)}{v'(C_0)} = \frac{1+\eta}{1+r} \tag{14}$$

Suppose that the interest rate happens to equal the rate of pure time preference:

$$r = \eta \tag{15}$$

In that case we have

$$v'(C_1) = v'(C_0) \tag{16}$$

If the function $v(\cdot)$ were convex from above everywhere, $v'(\cdot)$ would be monotonic decreasing, and the solution values would be

$$C_1 = C_0 = Y^* \tag{17}$$

the only point in (C_0, C_1) space that satisfies both (12) and (16). However, if $v(Y^*)$ is concave from above, as shown in Figure 1, the pertinent indifference curves, shown in Figure 2, are concave to the origin. Although at point A in Figure 2, representing (17), the necessary condition (16) is satisfied, the household prefers B, at which (16) is satisfied and at which U' of equation (13) is maximized. The household saves $(Y^* - C_0)$ currently to support high living at C_1 in the next period, where

$$C_1 - Y^* = (1+r)(Y^* - C_0) \tag{18}$$

Another way the household may be able to vary its consumption level is by gambling. Suppose that in each period fair gambles of any size and odds that the household wants are available, but that saving and dissaving are for the moment ruled out. With the utility function $v(\cdot)$ and income

---

[3] A general proof of concavity may be derived by following Hirshleifer (p. 233, fn. 24). Although Hirshleifer was dealing with a different economic problem involving the prospect of mutually exclusive alternatives, his mathematics are nonetheless the same as that needed to demonstrate concavity here.

Y^* shown in Figure 1, the household would choose a lottery ticket costing $Y^* - C_0$ having a payoff of $C_1 - C_0$, with a probability of payoff

$$p = \frac{Y^* - C_0}{C_1 - C_0} \tag{19}$$

The household would have the expected consumption level in each period

$$C_0 + p(C_1 - C_0) = C_0 + (Y^* - C_0) = Y^* \tag{20}$$

and its expected utility in the current period would be

$$E(v) = v(C_0)(1-p) + pv(C_1) \tag{21}$$

$$= \frac{(C_1 - Y^*)v(C_0) + (Y^* - C_0)v(C_1)}{C_1 - C_0} \tag{22}$$

Its expected utility for the two periods together, using (13), would be

(23)

$$U^* = \frac{(C_1 - Y^*)v(C_0) + (Y^* - C_0)v(C_1)}{C_1 - C_0} \cdot \left(1 + \frac{1}{1+\eta}\right) + V$$

$$= \frac{(C_1 - Y^*)v(C_0) + (Y^* - C_0)v(C_1)}{(C_1 - C_0)(1+\eta)} \cdot (2+\eta) + V \tag{24}$$

Now by (12) and (15) we have

$$\begin{aligned} C_1 - C_0 &= C_1 - Y^* + Y^* - C_0 \\ &= (Y^* - C_0)(1+r) + (Y^* - C_0) \\ &= (Y^* - C_0)(2+\eta) \end{aligned} \tag{25}$$

and hence

$$U' = U^* \tag{26}$$

The household has done as well by gambling as by saving: when $r = \eta$ the household is indifferent between the two.

Suppose next that we lower the rate of interest r for this household, letting other things remain the same. The outcome of gambling, U^*, is unaffected. But if the household chooses thrift, the budget constraint (12) will force the household to accept lower values of C_0 and C_1, and thereby a lower U'. Since U^* is greater than U', the household will prefer gambling to saving whenever $r < \eta$. Similarly, if we raise r above η, the household will prefer saving to gambling.

Note, however, that the household can normally also borrow in the period 0 to obtain the uneven pattern of consumption it wants, and a lower rate of interest will make this option more attractive. Thus, if r should be less than η for a particular household, and the household can borrow as well as lend at this rate r, it will find high spending out of borrowing now preferable to either thrift or gambling. Thus we arrive at one of our most important results: *if the rate r at which the household can lend and borrow is greater than η, the household will prefer saving to gambling at fair odds, whereas if it should be less than η, the household will prefer borrowing to gambling. Fair-odds gambling* (and a fortiori actual gambles) *will never be preferred*, provided that the household can save and dissave at the same interest rate. A necessary but not sufficient condition for a rational Friedman-Savage gamble is that there be either a rationing of the opportunities to borrow or save, or a difference in the borrowing and lending rates sufficiently great to exceed the cost of gambling due to unfair odds. Even when the marginal utility of consumption is rising, a Friedman-Savage gamble will not occur unless the cost of gambling (unfairness in odds due to transaction costs, taxation, or fear of laws against gambling) is less than the cost of saving at an interest rate below η and also less than the cost of borrowing at a rate above η. Thus the "actuarial loss' from gambling must be less than one-half the difference between borrowing and lending rates. (Business risks with an expected net gain, i.e., gambles at better than fair odds, may of course occur even where the investor is risk averse because of a diminishing marginal utility of income.)

By contrast, the hypothetical investor, imagined in countless articles, who is *a risk preferrer because of a rising marginal utility of income, cannot exist,* unless there are serious imperfections in the capital market. As we demonstrated in Section II, an individual confronting a rising segment in his utility function will react to any change in income within the relevant range by changing the amount of time spent at higher or lower levels of consumption—he will in this range have the constant marginal utility of consumption and risk neutrality depicted by the straight line from Q to R on Figure 1. For any changes outside the QR interval the individual will be risk averse. Only if the household is constrained in saving and dissaving could it prefer risk on Friedman-Savage grounds; it might then demonstrate this risk preference by accepting unfair odds if fair gambles were also unavailable. Risk-averse behavior can plausibly be explained by a declining marginal utility of income, but any risk preference can be explained only in terms of a taste for adventure or faulty capital markets.

V. Nonseparable Utility

What difference does it make when we drop the assumption of separable utility? There certainly can be temporal interdependence in utility functions, as when consumption falls below the survival level and thereby precludes future consumption. Whether survival is at issue or not, households may prefer relatively even levels of consumption; the memory of a period of lavish consumption might make periods of low consumption intolerable, or the prospect of austerity might spoil periods of high living, or experiences of low consumption might make people ill at ease in later luxury. Indeed, it is conceivable that such a preference for an even consumption pattern rather than diminishing marginal utility explains the observed stability of consumption behavior. With nonseparability there is also the logical possibility that the indifference curve in Figure 2 might be convex while the function relating utility to lifetime consumption (or wealth) nevertheless had a positive second derivative. Consider, for example, a utility function that over a range has the form

$$U = C_0 \cdot \frac{C_1}{1+\eta} \tag{27}$$

If $r=\eta$ the household saves or dissaves to equalize consumption in the two periods. Yet a doubling of consumption quadruples utility.

We have thought of a possible real counterpart to utility functions that would have a range of this kind. Suppose that a move to a higher income and social class would bring a disproportionate increase in utility, but that some of the associations or other benefits of the upward mobility could be attained only if a high and stable level of consumption is maintained for a very long time. It might not be possible to achieve a sufficient number of successive periods of high consumption by saving or borrowing, but they might be attained with good luck on a big gamble. The individual would not (unless he received windfall income) gamble more than once, since a single gamble at appropriate odds will leave him at an "efficient" level of consumption and a further gamble whether won or lost will work against the postulated desire for extended stability in consumption levels.

Are the particular patterns of temporal interdependence that might permit both the stable consumption levels that are observed and also a rising marginal utility of income empirically significant? And are the gambles that are observed in reality consistent with these nonseparable utility functions? We have not undertaken an exhaustive analysis of these questions and must leave them unsettled. But we doubt that rising marginal utility is commonplace and that much actual gambling is explained on Friedman-Savage lines.

First, if there were a rising marginal utility of lifetime consumption or wealth, the rational gambler would normally, as in our example, make a single large gamble at appropriate odds. Continued gambling, moreover, would run counter to the purpose of avoiding a range of rising marginal utility of income; a large number of gambles would tend to average out, leaving the gambler close to the actuarially average outcome. Thus the observed behavior of gamblers, who typically gamble repeatedly, cannot be explained on Friedman-Savage lines even with the aid of a helpful nonseparable utility function like (27). Indeed, typical gambling behavior suggests the alternative hypotheses that Friedman and Savage reject: that there is utility in the gamble itself, or gamblers fail to recognize the odds.

This leads to the second ground for skepticism. Friedman and Savage argue that lotteries, at least when the purchaser of a lottery ticket is not a spectator to the drawing of the winners, "present risk in relatively pure form, with little admixture of other factors" (p. 287), such as the entertainment value of games of chance, which they suppose could be sold separately from the monetary risks and rewards and played "in a game using valueless chips." In fact, enough data have now accumulated from state lotteries to leave little doubt that the participants are buying entertainment, or making what Adam Smith called "absurd presumptions about their own good fortune," rather than revealing anything about the curvature of their utility functions. It appears (see Tom Stevenson) that:

1. Most adults in lottery states have purchased at least one ticket. About 45 percent buy at least one a month, and 30 percent buy at least one or two a week.
2. Usually only 40 to 50 percent of the receipts from the lottery tickets is paid back in prizes.
3. Participation in a given lottery game variation drops after a few weeks or months, so that the typical state runs two or three new games per year.
4. The most appealing prize structure has been found to be to give out about two-thirds of the prize money in large numbers of small prizes of $2 to $5, and giving out the balance in a few very large prizes—$10,000 or higher. The many small ones are said to inspire confidence, the few large ones to attract publicity.
5. Detailed features of the successful lotteries indicate that it pays the promoter to make a game out of it.

The last three points in particular indicate that the lotteries are adventurous entertainment or designed to mislead participants about what the true odds are. The fourth point also suggests a pattern of prizes that contradicts Friedman and Savage's prediction that the range of prizes offered will be designed to attract gambles by individuals with concave-upward segments of different lengths in their utility functions.

We therefore conclude that Friedman's and Savage's interpretation of gambling is wrong, as is Friedman's related attempt to explain income distributions in terms of the curvature of utility functions. But their seminal contributions also illustrate Francis Bacon's point that "truth will sooner arise from error than from confusion." As a result of the clear and provocative line of thought which they helped begin, we can now say that risk preference due to a rising marginal utility of income could occur, if at all, only in remarkable conditions. We now also have a basis, in the observed stability of consumption levels and the lack of Friedman-Savage gambles, for the conclusion that the marginal utility of income characteristically decreases with income.

REFERENCES

M. J. Bailey and M. Olson, "Pure Time Preference, Revealed Marginal Utility, and Friedman-Savage Gambles," work. paper no. 77-2, dept. of econ. and Bur. Bus. Econ. Res., Univ. Maryland 1977.

G. Debreu; "Topological Methods in Cardinal Utility Theory," in Kenneth Arrow et al., eds., *Mathematical Methods in the Social Sciences*, Stanford 1959.

M. Friedman, "Choice, Chance, and the Personal Distribution of Income," *J. Polit. Econ.*, Aug. 1953, *61*, 277–90; reprinted in *Price Theory*, Chicago 1976.

______ **and L. J. Savage,** "The Utility Analysis of Choices Involving Risk," *J. Polit. Econ.*, Aug. 1948, *56*, 179–304; reprinted in *Readings in Price Theory,* Homewood 1952.

Jack Hirshleifer, *Investment, Interest, and Capital*, Englewood Cliffs 1970.

T. C. Koopmans, "Stationary Ordinal Utility and Impatience," *Econometrica*, Apr. 1960, *28*, 287–309.

M. Olson and M. J. Bailey, "Revealed Utility Functions and Friedman-Savage Gambles," mimeo. 1978.

——— **and** ———, "Positive Time Preference," *J. Polit. Econ.*, forthcoming.

P. A. Samuelson, "A Note on Measurement of Utility," *Rev. Econ. Stud.*, Feb. 1937, *4*, 155–61.

Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, London 1776, Book 1, ch. X.

T. Stevenson, "Mastermind of the Instant Lottery," *New York Times*, Jan. 2, 1977.

R. H. Strotz, "Myopia and Inconsistency in Dynamic Utility Maximization," *Rev. Econ. Stud.*, No. 3, 1956, *23*, 165–80.

P. Wonnacott, "The Permanent Income Hypothesis and the Friedman-Savage Utility Function," mimeo. 1972.

On the Method of Taxation and the Provision of Local Public Goods

By DAVID A. STARRETT*

Will independent communities tend to oversupply or undersupply local public goods? This question has been debated in the literature extensively since it was first voiced by Charles Tiebout. Persuasive arguments can be given on both sides of the issue. A typical argument for undersupply would run as follows: Local communities will tend to discourage expansionary projects because such projects will encourage in-migration and increase congestion to current residents. However, the increase in congestion in a single community is not a real social cost, since congestion levels in other communities will fall as a result of the migration. Hence, local communities will be too restrictive in their provision of local public services.

On the other hand, a contrasting argument for oversupply can be given. Communities, in attracting residents, can bid tax revenue away from others; the extra tax revenue can be used to spread the costs of a given provision of public goods more thinly over the original population. Since the additional tax revenue is a pure transfer among communities, there is a private benefit to the expanding community, but no corresponding social benefit. Hence, individual communities will attempt to overexpand (and will succeed to the extent that they can attract residents).

Of course, there is a third view, that "voting with one's feet" will tend to generate a first best distribution of local public goods; indeed, this is the view most generally attributed to Tiebout. A recent exponent of this view is Bruce Hamilton. He argues that, given a sufficient number of communities, the efficient Tiebout equilibrium can be enforced through the offering of a spectrum of public service/tax packages; the taxes can be either lump sum in nature or property taxes (with zoning restrictions to eliminate the possibility of free riders). However, the Hamilton model is not immune to competition among communities. Even if all residents pay their fair "share" of the costs of public goods provision, a community can still benefit by attracting an extra resident so as to lower the cost of public services to all current residents. It is precisely these incentives which I wish to study.

It is worth noting one other attack on the Tiebout "efficiency" theorem, although I will not be dealing with it directly in the sequel. In a series of papers by James Buchanan and Charles Goetz, and Frank Flatters, Vernon Henderson, and Peter Mieszkowski, all argue that free mobility will not generate a Tiebout optimum. These authors note that there may be a misallocation in the migration mechanism because migrants ignore fiscal externalities in the communities they leave. They analyze the resulting distortions for some simple models of location. However, it is difficult to relate these results to the general Tiebout hypothesis since Buchanan and Goetz, and Flatters, Henderson, and Mieszkowski explicitly restrict the range of choice available to consumers (in the latter paper, there are only two communities from which to choose). Tiebout and his followers would surely agree that restrictions in the range of choice will lead to distortions. By contrast, we will be concerned here with distortions induced by gaming behavior on the part of communities.

In order to study the incentive question rigorously, let us first construct a welfare formulation of the problem in which the statements "too large" and "too small" have precise meaning. Having done this, we can

---

*Stanford University. This work was supported by the National Science Foundation Grant SOC77-06000 at the Institute for Mathematical Studies in the Social Sciences, Stanford University, and in part by a grant from the Guggenheim Foundation.

compare the private incentive to expand with the corresponding social incentive under a variety of different taxation and zoning schemes.

The main conclusions of my inquiry can be summarized as follows: First, standard cost-benefit analysis gives somewhat misleading answers to questions concerning optimal dispersion of economic activity, due to a subtle "second best" feature of the problem studied. I identify this feature and discuss its implications. Second, I show that each of the intuitive arguments given above can be supported under particular types of taxation schemes. Communities will try to overexpand if the local tax is on sales or wage income, or the tax is a property tax which is imposed on "improved" land only. Communities will try to "underprovide" if the tax is a property tax on all land (improved and unimproved).

I. The Prototype Model

I will employ a simple prototype model of local communities in this paper. Some of the conclusions of this paper remain valid in a much more general framework; that work is reported in my forthcoming paper.

Each community provides a single local public good (q). This good is public in the sense that access to it is without direct charge, although residents may pay something (for example, transport costs) to gain access, and the public facilities may become congested with use. Museums, civic centers, parks, and libraries all fit reasonably well into this framework.[1] In the following discussion I focus on a particular community (k).

Aside from land, it is assumed that there is only one other private good. (Alternatively, it could be assumed that other private goods prices are fixed throughout the analysis and these goods could be aggregated using the fixed prices as weights.) Let this private good be the numeraire.

As mentioned above, agents may have to engage in effort (take trips) to enjoy the public good. Assume that the expense of such trips can be represented by a simple cost function in terms of the numeraire ($f^k(g,s)$), where g is the number of trips taken, and s stands for location within the community. Location will be specified by dividing up the community land into discrete zones, and letting s be a zoning index.[2]

Let us also assume that consumers can be divided up into a finite number of types so that all members of a given type have the same preferences and endowments. The letter a will be used frequently to index consumer type.

The value of endowment may depend on the community in which it is held (since prices and production opportunities may differ among communities). For convenience, assume that agents own resources initially in only one community. Let j index the community of ownership.

Each consumer (living in community k) gets to choose a zone (s) and an amount of land within the zone (l_s), in addition to the choice of g. The market for land will be assumed competitive and r_s^k will stand for the (net of tax) rental rate in zone s. Consumers earn income in terms of the numeraire. We need to separate this income into three terms: profit shares (P), rental income on land (R), and other income (mainly labor) (Y). Taxes will be assessed to pay for the public goods; at this stage I model these in a fairly general way to consider several types of taxation. Potentially, the taxes paid by an a-type agent depend on his decisions regarding location and land use, as well as tax parameters t_j^{ask}, i.e., $T = T(l,s,t_j^{ask})$. Assume, however, that the taxes do *not* depend *directly* on the use of the public facilities (as measured by g).

Holding community of residence (k) fixed for the moment, a representative a-type agent owning property in j solves the problem:[3]

$$(1)\quad \operatorname*{Max}_{g,l,s} U^a\big(g,l,q^k,Q^k,Y^{ak}+R_j^a + P_j^a - f^k(g,s) - r_s^k l - T(l,s,t_j^{ask})\big)$$

---

[1] My framework is similar to that employed by Frank Westhoff except that I drop the assumption of radial symmetry.

[2] Implicitly, I am assuming that land within each zone is perfectly homogeneous in all "economic" respects.

[3] Since labor resources move with the owner, other income (Y) will depend on the community of residence.

where U^a stands for a representation of a-type preferences, and Q^k is the congestion level, which is treated as a parameter by the individual. It seems appropriate that both g and q should affect utility. For example, if the public good is a museum, households will care about the size of the museum as well as the number of visits made. Note that "left-over" income is automatically consumed so there is no possibility of saving (the model is static in nature). Finally, observe that we are implicitly assuming free mobility within the community since zone (s) is a free choice.

It is convenient to think of the choice of zone as being made "last." Let us define an indirect utility function conditional on s in the usual way:

$$(2) \quad V^a\left(q^k, Q^k, Y^{ak} + R_j^a + P_j^a, r_s^k, s, t_j^{ask}\right) = \operatorname*{Max}_{g,l} U^a$$

Frequently, I will indicate the indirect utility function in (2) using the shorthand notation V_j^{ask}. And the associated choice variables will be indicated by l_j^{ask}, g_j^{ask}.

The ultimate indirect utility function of our agent (V_j^{ak}) is then defined by the "free choice" problem:

$$V_j^{ak} \equiv \operatorname*{Max}_{s} V_j^{ask}$$

I denote by n_j^{ask} the number of such agents who choose zone s.

I treat production in the community in a completely standard way. Each production unit behaves competitively, and none are affected by the public good. Profit maximization generates a profit function for each unit; the arguments of these functions are private goods prices, and since the nonland private good is the numeraire, the profits are functions of rental rates only. Since firms are final users of land, the rental rates they pay must be *gross* of any property tax. Aggregating across firms, we can derive an aggregate profit function of the form $\Pi^k = \Pi^k(\rho)$, where ρ stands for the gross-of-tax rent vector.

The sole function of local government is to produce public goods from private goods, paying for those goods with tax revenue. Letting $\Gamma^k(q,\rho)$ stand for the public goods cost function in k, the government must satisfy a balanced budget condition:

$$(3) \quad T^k = \sum_{askj} T_j^{ask} n_j^{ask} = \Gamma^k(q^k, \rho^k)$$

This completes the basic structure of a community although I will need to discuss the equilibrium determination of rents and congestion levels later on. Notice that public goods are providing the only economies of scale to community size, since my neoclassical treatment of the firm precludes increasing returns to scale in private production.

Consumers have one final choice to make: the choice of residence community. They face a Tiebout-type selection of tax/public service packages. Here, I will take the extreme position that there is free migration *among* as well as *within* communities, enabling us to study migration effects in their pure form. Clearly, the migration influence is not a short-run phenomenon and the analysis to follow should be interpreted with that in mind. "Free" choice of community means that each agent maximizes indirect utility over choice of k so that the utility achieved by an agent can depend only on his or her type and place of ownership; it cannot depend on where he or she lives. The "realized" utility for an a-type person owning in j is defined by

$$V_j^a \equiv \operatorname*{Max}_{k} V_j^{ak} \qquad \text{each } a,j$$

Communities potentially may interact with each other in a number of different ways. Here we will assume away some of the potential interactions in order to concentrate on the issue of migration. First, let us assume away direct spillover effects; that is, assume that residents of one community get no benefits from public goods elsewhere. This assumption is already implicit in the formulation of the consumer problem (1) since only q^k and Q^k were allowed to affect the utility of a k resident.[4] Second, let us assume away any interactions through trade. This does not mean that we rule out trade altogether; indeed, we allow free trade among communities.[5] However, we have assumed that the prices of tradeable goods do not change as a result of any proposed projects, so that projects in any one community cannot change the trade opportunities of any other community. Third, let us rule out any interactions through absentee ownership of land or firms (which use land). This is done by prescribing local ownership *ex ante* ($n_j^{ask}=0, j\neq k$).[6] Local ownership in land means

$$\sum_{as} n_k^{ask} R_k^a \equiv R^k \qquad \text{each } k \tag{4}$$

where R^k is the value of all land in community k, while local ownership of firms requires:

$$\Pi^k = \sum_{as} n_k^{ask} p_k^a \qquad \text{all } k \tag{5}$$

where Π^k is the value of all profits generated in k.

I will relax the ownership assumption somewhat later in this paper. However, the reader is referred to my forthcoming paper for analysis of a general model without these assumptions.

The only remaining interaction among communities is through migration. When one community undertakes a project to improve the welfare of its residents, this action may be expected to attract residents from other communities. We are not in a position to derive the migration response function analytically without more information, but we can derive measures of incentive bias as a function of induced migrations.

At the outset, consider the model in a "status quo ante" position (before any new projects are considered). In this position, each community will be providing a certain level of public goods (not necessarily optimal), and setting tax rates in such a way as to balance its budget. For any such configuration, it is assumed that there exists a private sector equilibrium which is characterized by two conditions. First, private goods prices must adjust to clear all markets; here, this means that the rental rates on land (in terms of the numeraire) adjust to clear the market for land at each location in each community. Second, the congestion levels must be self-confirming (in the Nash sense). Here, we take the simple view that congestion levels can be measured by total use of the facility. Hence, the Nash equilibrium conditions will be

$$Q^k = \sum_{as} n_k^{ask} g_k^{ask} \qquad \text{all } k$$

The reader is referred to my forthcoming paper for a discussion of conditions guaranteeing private sector equilibrium.

II. The Welfare Formulation

Let us measure welfare in the standard way that lies behind cost-benefit analysis.[7] It is assumed that the governments control public goods provision and tax rates, but cannot interfere with private markets (or individual choice) in any other way. Further, we require that social preferences respect private preferences. Therefore, the social planners must take *indirect* utility functions as arguments in their social welfare function. In resolving the issue of interpersonal comparisons, it is assumed that a minimal horizontal equity condition is satisfied: any two agents of the same type must be treated symmetrically in the sense that interchanging them in the arguments of the welfare function does not change the social ordering. Since (under our assumption of free mobility) all agents of a given type who own in a particular community are sure to get the same utility, symmetry implies that all "similar" agents must have the same marginal welfare weight in any initial status quo position.

---

[4] Thus our analysis is best suited to the study of communities separated in space. For a general discussion of direct spillover, see Buchanan (1968).

[5] Implicitly it is assumed that trading activities are carried out by firms in the various communities.

[6] Local ownership will be destroyed as soon as new migration takes place. But in our first-order analysis, this will not matter.

[7] For a good exposition of this theory, see Robert Willig.

I will confine myself here to statements about the first-order welfare impact of a project.[8] Given the welfare formulation outlined above, it is easy to show that the first-order welfare change can be written in the form

$$dW=\sum_{aj}\omega_j^a\left(\sum_{s,k}n_j^{ask}\right)dV_j^a$$

where ω_j^a is the marginal welfare weight for a-type agents, owning initially in j.

The next step in formulating cost-benefit analysis is to assume that the initial distribution of income is optimal (so that the marginal welfare of a dollar is the same wherever applied). However, even if we believed that social planners were committed to such a policy, optimality cannot be achieved here because of the inherent indivisibility in location choice. If two people of the same characteristics live at different places within a given community, then (except by chance) they will not have the same marginal utility of an extra dollar; indeed, we might expect that the agent living further out would have a lower marginal utility since he or she would have to use some of the dollar for transport. Thus, our two agents *cannot* have the same marginal welfare of a dollar. This argument is essentially that of James Mirrlees and is an illustration of his general proposition that identical agents must be given different incomes in the first best optimal town.

Since we cannot optimize on the distribution of income, the welfare formulation (as I have given it) is inherently second best, and this fact has important implications when we come to the interpretation of results. Further, the presence of a second best distortion ordinarily will lead to a second best term in first-order welfare measures. However, this is not the case here as long as incomes are distributed optimally among types (on average). Intuitively, all projects must have a neutral effect on the distribution of utilities within type (leaving it flat) so that projects can neither improve nor worsen the second best position. Analytically, this neutrality means that the first-order welfare effect still can be measured in terms of dollar aggregates just as it is whenever incomes are optimally distributed. Formally, let λ_j^{ask} stand for the marginal utility of the numeraire in problem (2). Then if incomes among types are adjusted so that

$$\omega_j^a=\sum_{ks}\left[\frac{n_j^{ask}}{\sum_{\sigma,\kappa}n^{a\sigma\kappa}}\right]\frac{1}{\lambda_j^{ask}}$$

welfare change can be written in the form

$$dW=\sum_{askj}n_j^{ask}\frac{dV_j^{ask}}{\lambda_j^{ask}}$$

Note that income is distributed in such a way that the reciprocal of the welfare weight (for each type) is equal to the harmonic mean of the marginal utilities of a dollar (for that type).

It is important to stress at this point that I am not defending the assumptions of cost-benefit analysis here; I am merely exploring some of their implications in local public finance. Naturally, if the income distribution is not optimal then the subsequent analysis must be modified to take into account distributional aspects of projects' benefits. Perhaps this could be done through the use of welfare weights, although the possibility is not explored here.[9]

---

[8] Integrating the expression for first-order change is not legitimate since some of the variables (for example, migration variables) are discrete. For further discussions of this point, see my forthcoming paper.

[9] See Burton Weisbrod for a general discussion of the use of welfare weights.

III. Direct Taxation on Residents

Taxes will be of the lump sum variety in this section. Since no labor-leisure or consumption-saving margin is allowed, we could equally well think of the taxes as being on income or consumption. Thus, $T_j^{ask} = t_j^{ask}$, all a,k. Let us now focus on a particular community (z) and consider a new project in that community (dq^z). This project naturally induces changes in the other parameters faced by our economic agents, and it is assumed that these parameters react in a continuous way.[10] We seek to measure the net welfare benefit from such a project both from the point of view of society and from that of community z. Intuitively, any gap between these measures will correspond to a bias in community incentives.

To compute the change in social welfare, first compute the change in individual welfare and then aggregate. Totally differentiating (2) and utilizing standard duality results yields[11, 12]

$$(6)\quad \frac{dV_j^{ask}}{\lambda_j^{ask}} = \frac{\dfrac{\partial U_j^{ask}}{\partial q^k}}{\lambda_j^{ask}} dq^k + \frac{\dfrac{\partial U_j^{ask}}{\partial Q^k}}{\lambda_j^{ask}} dQ^k - l_j^{ask} dr_s^k + dP_j^a + dR_j^a - dT_j^{ask}$$

Aggregating, substituting for profit income from (5), and applying duality theory to evaluate the differential of profit, yields

$$(7)\quad dW = \alpha_q^z dq^z - \sum_k \alpha_Q^k dQ^k - \sum_k \left(L_r^k + L_f^k\right)\cdot\delta r^k + \sum_k dR^k - \sum_{ask} n_k^{ask} dT_k^{ask}$$

where α_q^k stands for the direct marginal social benefit per unit of q in k, α_Q^k stands for the corresponding marginal social cost of an extra unit of congestion in k, L_r^k stands for the vector (over zones) of land put to residential use in k, and L_f^k stands for the corresponding vector of land put to industrial use in k.[13]

The market parameters (rents, taxes, etc.) depend on project values through market equilibrium and budget balance conditions. We can eliminate some of those parameters, and simplify the expression for dW considerably by making use of those conditions (in differential form). Equilibrium in the market for land together with the ownership conditions (4) implies

$$(8)\quad dR^k = \left(L_r^k + L_f^k + L_g^k\right)\cdot\delta r^k \equiv L^k\cdot\delta r^k$$

where L_g^k stands for the vector of land used in the public goods project and L^k is the fixed vector of total land available in k. Since the government must pay for a new project out of tax revenue,

$$d\Gamma^k = \sum_{asj} n_j^{ask} dT_j^{ask} + \sum_{asj} T_j^{ask} dn_j^{ask}$$

And, applying duality theory to the government cost function:

$$d\Gamma^k = -L_g^k\cdot\delta r^k + \frac{\partial\Gamma^k}{\partial q}\cdot dq^k$$

After substituting all these equilibrium relationships into (7) we derive finally

$$(9)\quad dW = \alpha_q^z dq^z - \frac{\partial\Gamma^z}{\partial q} dq^z + \sum_{askj} T_j^{ask} dn_j^{ask} - \sum_k \alpha_Q^k dQ^k$$

Before discussing the private welfare incentive, let us analyze the form of (9). The

---

[10] Changes in land rent are expected to prevent any unstable swings in migration.

[11] Assume that all functions are sufficiently smooth so that the standard envelope theorems are applicable. For an exposition of the theory of duality as applied here, see W. Erwin Diewert.

[12] $dY^{ak} \equiv 0$ due to the assumption that nonland private goods prices are fixed.

[13]

$$\alpha_q^k = \sum_{asj} n_j^{ask} \frac{\partial U_j^{ask}}{\partial q^k} + \lambda_j^{ask}; \alpha_Q^k = -\sum_{asj} n_j^{ask} \frac{\partial U_j^{ask}}{\partial q^k} + \lambda_j^{ask}$$

$$L_s^k = \sum_{asj} n_j^{ask} l_j^{ask}$$

first two terms reflect the standard "first best" contributions to consumer's-plus-producer's surplus. It is worth noting that the marginal costs of the project are to be computed at fixed prices; any increases in costs due to increases in prices cancel out against the corresponding net benefit to suppliers.

There are two kinds of second best terms which modify the first best measure. The first type measures the net change in "costs" of congestion externalities and needs little comment. The second type measures the net increase in tax revenue to society from migration,[14] and needs some discussion since we are not used to thinking of tax revenue as measuring net economic benefits to society. The reason that a tax term appears here is that the Tiebout-type tax system (with different communities setting different lump sum taxes) is not a true lump sum tax system as long as individuals are free to move; the tax one pays is a function of the decision as to where to live.

Consequently, any project which induces migration must have a second best welfare impact as long as tax rates differ among communities. Indeed, when an individual voluntarily moves from a low-tax area to a high-tax area, he or she confers more tax-spreading benefits on the recipient community than he or she removes from the original community. The net increase in tax revenue is a real social benefit.[15]

Let us now ask what the benefits look like from the point of view of community z. Before answering this question, it must be formulated more precisely. In particular, who are residents? The position will be taken here that residents are those households who are in the community *before* the project is initiated. There are several arguments for this position. First, current residents constitute the constituency which will decide on the project. Thus, if we want external effect to have its usual meaning of "net benefits or costs to those outside the decision making group," the outsiders should be those outside the community before the project is initiated.

Second, if benefits are measured relative to the group of residents present *after* the project is initiated, there will be an obvious direct incentive for every community to try to attract rich people (who will raise the average level of welfare in the community). I doubt that community objective functions have that property. Of course, even with our definition of resident, there is still an *indirect* incentive to attract rich people, if they can be made to pay a larger share of taxes than others. This incentive seems to me real, and I want to measure the possible distortions it will introduce into decision making.

Let dW^z stand for the welfare benefit from the project to initial residents of community z. Performing exactly the same sequence of steps as above, we derive

$$dW^z = \left(\alpha_q^z - \frac{\partial \Gamma^z}{\partial q}\right) dq^z + \sum_{asj} T_j^{asz} dn_j^{asz} - \alpha_Q^z dQ^z$$

The gap between social and private incentives (dW^0) is just the difference between dW and dW^z:

$$dW^0 = \sum_{\substack{k\neq z\\ asj}} \left[T_j^{ask} dn_j^{ask} - \alpha_Q^k dQ^k \right]$$

This expression can be simplified somewhat; recall that the only effect on congestion in communities other than z is through migration. Therefore,

$$dQ^k = \sum_{asj} g_j^{ask} dn_j^{ask} \qquad k \neq z$$

from which we can infer

$$(10) \qquad dW^0 = \sum_{\substack{k\neq z\\ asj}} \left(T_j^{ask} - \alpha_Q^k g_j^{ask} \right) dn_j^{ask}$$

How should we interpret (10)? If communities other than z tax at rates generally higher than their marginal congestion cost, and the project in z induces migration toward z, then (10) tells us that the private incentive to "expand" exceeds the social incentive, and vice versa. Notice that if all communities set the tax rate (on average) equal to marginal congestion costs, the incentive gap disappears, and indeed, all second best migration terms disappear from our welfare measures. This observation suggests that a congestion tax may be "desirable."

---

[14] dn_j^{ask} is positive if (a,j) agents move into (s,k) but negative if they move out. Each individual who moves into one community must have moved out of some other community, so the third term in (9) measures the *net* changes in tax payments.

[15] It does not follow that high-tax communities are a good thing per se. For one thing, high-tax communities may tend to be especially congested in which case the tax term may be more than offset by the congestion term. But more importantly, the analysis applies only to *voluntary* moves; forcing people to move from low-tax areas to high-tax areas certainly need not improve welfare.

A discussion of the desirability of congestion taxes will be postponed until later. First I want to rework the analysis with property taxes in place of lump sum taxes. Surprisingly (perhaps) the nature of the conclusions changes completely.

IV. Property Taxation with Blind Zoning

I will consider only the case of pure ad valorem property taxes and let t^k stand for the tax rate in community k.[16] In this section it will be assumed that the amount of land subject to taxation in each community is fixed. A more flexible zoning arrangement is considered in the next section.

Let us formulate the tax system as if the taxes are imposed by statute on users of the land, although this formulation is inessential to the argument since statutory incidence is of no importance as long as the market for land is assumed competitive. Thus, if an agent uses an amount of land l in zone s of community k, he or she will pay tax $t^k r_s^k l$, so

$$T_j^{ask} = t^k r_s^k l_j^{ask} \qquad \text{all } a,s,j,k$$

Recomputing (6) with this tax specification, we have

$$(11)\quad \frac{dV_j^{ask}}{\lambda_j^{ask}} = \frac{\partial U_j^{ask}}{\dfrac{\partial dq}{\lambda_j^{ask}}} dq^k + \frac{\partial U_j^{ask}}{\dfrac{\partial Q^k}{\lambda_j^{ask}}} dQ^k$$
$$+ dP_j^a + dR_j^a - r_s^k l_j^{ask} dt^k - (1+t^k) l_j^{ask} dr_s^k$$

Now we carry out the steps of the previous section. In so doing there are two slight modifications. First, we must recognize that firms, as users of land, pay gross-of-tax rental rates. Therefore,

$$(12)\quad d\Pi^k = -L_f^k \delta\rho^k = -r^k \cdot L_f^k dt^k$$
$$-(1+t^k) L_f^k \cdot \delta r^k$$

Aggregating over all agents in (11) and substituting from (12) yields the intermediate formula

$$(13)\quad dW = \alpha_q^k dq^k - \alpha_Q^k dQ^k + dR^k$$
$$-(1+t^k)\left(L_f^k + L_r^k\right) \cdot \delta r^k$$
$$-r^k \cdot \left(L_r^k + L_f^k\right) dt^k$$

Second, the form of the government budget balance equation is different:

$$(14)\quad t^k r^k \cdot L^k = t^k r^k \cdot \left(L_r^k + L_f^k + L_g^k\right)$$
$$= \Gamma(q, r^k(1+t^k))$$

Note that the government also pays gross-of-tax rents for the land it uses in public projects. Differentiating (14) totally and applying duality theory to the government cost function, yields

$$(15)\quad \left(L_r^k + L_f^k + L_g^k\right) \cdot \left[t^k \delta r^k + r^k dt^k\right]$$
$$= \frac{\partial \Gamma}{\partial q} dq + L_g^k \cdot \left(r^k dt^k + (1+t^k)\delta r^k\right)$$

Finally, substituting (15) and (8) into (13) and performing some cancellations, we derive

$$(16)\quad dW = \alpha_q^z dq^z - \sum_k \alpha_Q^k dQ^k - \frac{\partial \Gamma^z}{\partial q} dq^z$$

Observe that the formula is exactly as before except that the tax terms have disappeared.[17] And it is easy to show that the other formulas are modified in exactly the same way; i.e.,

---

[16] Of course, much of the flexibility in redistributing income is given up once this type of tax is adopted. However, it is not necessary that local taxation serve to distribute income optimally as long as there is an appropriate national policy determining the "post-redistribution" Ys and Rs.

[17] In my static model (without capital goods) there are no second best distortions due to taxation of land.

$$(17)\quad dW^z = \alpha_q^z dq^z - \alpha_Q^z dQ^z - \frac{\partial \Gamma^z}{\partial q} dq^z$$

$$(18)\quad dW^0 = -\sum_{k \neq z} \alpha_Q^k \left(\sum_{asj} g_j^{ask} dn_j^{ask} \right)$$

The tax terms disappear in all these measures because there is no way of changing the tax base in any of the communities. When one individual moves into community z from outside, he or she does indeed pay some taxes. However the individual(s) who are displaced pay less taxes by the same amount (unless the rental rates or tax rates have changed). But changes in the rental and tax rates have no net welfare implications per se; an increase in the tax rate helps the government but hurts land users by an equal and opposite amount. Also, changes in the rental rates which leave tax rates unchanged have a neutral net effect; owners gain from an increase but users lose, and vice versa. Only changes in the tax *base* can have net welfare implications.

The above statements are true both from the point of view of society and from the point of view of community z, since the owners and users of land in each community are all initially in that community. Absentee ownership naturally will modify the results from the point of view of community z (though not from the point of view of society). Let us analyze the implications of (16)–(18). The gap between private and social incentive is now equal to the change in congestion cost induced in other communities. Thus, communities will always be biased against expansion as long as there is any congestion at all in other communities. Only if communities were small enough, so that there was no congestion at all, would private and social incentives be in agreement.

How can it be that changing the nature of taxation changes the character of the results so completely? The answer is that we have changed the social opportunity set in an important way; once the tax base is fixed, we have restricted the "social flexibility" in choosing the size of communities. Indeed, it would be very surprising if the second best solution with fixed tax base had any implications at all for the first best dispersion of economic activity, and it will be shown that it does not in the next section. However, it is still true that *if* (for whatever political reason) we constrain ourselves to a world in which communities employ property taxes on a fixed tax base, then individual communities will find private incentives to be "more exclusive" than is desirable.

V. Property Taxation with Agricultural Zoning

Now consider a case of property taxation in which the tax base can be expanded in response to market forces. To this end, think of the community land as being divided up between improved and unimproved land, with only the improved land being taxed. Unimproved land may be agricultural land or it may simply be unincorporated land. However, let us take the position that any boundary between these two types of land is subject to change in response to market forces; that is, the zoning of a particular piece of land is determined by the use to which the highest bidder for that land wants to put it. Hence there must be disadvantages associated with using unimproved land; either there must be building restrictions, use restrictions, or prohibitions on the use of public facilities for those occupying unimproved land.

We can incorporate the type of zoning described above into the notational framework of this paper most easily if we are willing to make one further simplifying assumption. Let us suppose that agents are divided (by type) into two mutually exclusive classes, one of which always uses improved land and one of which always uses unimproved land. For households, let A^i stand for the class of types who use improved land and A^u for the class who use unimproved land. Then, for $a \in A^i$, the ad valorem property tax must be paid on any land used, so we have $t^{ak} = t^k$, and

$$T_j^{ask} = t^k r_s^k l_j^{ask}, a \in A^i \qquad \text{all } s, j, k$$

On the other hand for $a \in A^u$, $t^{ak}=0$, since no taxes are imposed on unimproved land.

Similarly, firms are divided into two classes. Let Π^{ki} denote the profit function of firms using improved land (and therefore paying taxes), and Π^{ku} denote the profit function of firms using unimproved land. The constraint on total land in community k is now written in the form

$$(19)\quad L^k = L^{ki} + L^{ku}$$

$$= L_r^{ki} + L_f^{ki} + L_g^{ki} + L_r^{ku} + L_f^{ku}$$

where L_r^{ki} stands for the vector of all land used for residential purposes by the class A^i, and the other terms have analogous interpretations. It is assumed that free-market forces will establish (19) as a market-clearing condition; in particular, the division of land between improved and unimproved uses is determined by competition between the associated groups.

We can now compute the welfare change to residents of community z. Aggregating separately over the initial residents from classes A^i and A^u, combining results, and using (5), we derive

$$dW^z = \alpha_q^z dq^z - \alpha_Q^z dQ^z + dR^z + d\Pi^z - L_r^{zi}\cdot\delta r^z$$

$$- t^z L_r^{zi}\cdot\delta r^z - r^z\cdot L_r^{zi} dt^z - L_r^{zu}\cdot\delta r^z$$

Next, profits can be broken into its two component parts, yielding

$$d\Pi^z = d\Pi^{zi} + d\Pi^{zu} = -L^{zi}\cdot\delta r^z - t^z L_f^{zi}\cdot\delta r^z$$

$$- r^z\cdot L_f^{zi} dt^z - L_f^{zu}\cdot\delta r^z$$

The final component needed in our analysis (the government budget balance condition) is of the same form as before, except that now we must take account of the fact that the tax base may change as a result of the project (for example, the improvement may attract outside taxpayers). Such a change leads to a first-order contribution to tax revenue equal to $t^z r^z\cdot\delta L^{zi}$. The budget balance condition in differential form becomes

$$\frac{\partial\Gamma^z}{\partial q}\cdot dq^z = L^{zi}\cdot r^z dt^z + t^z L^{zi}\cdot\delta r^z + t^z r^z\cdot\delta L^{zi}$$

$$- L_g^{zi}\cdot r^z dt^z - t^z L_g^{zi}\cdot\delta r^z - L_g^z\cdot\delta r^z$$

$$= \left(L_r^{zi} + L_f^{zi}\right)\cdot r^z dt^z + t^z\left(L_r^{zi} + L_f^{zi}\right)\cdot\delta r^z$$

$$- L_g^z\cdot\delta r^z + t^z r^z\cdot\delta L^{zi}$$

Performing the same sequence of substitutions as in previous sections, we obtain finally

$$(20)\quad dW^z = \left[\alpha_q^z - \frac{\partial\Gamma^z}{\partial q}\right] dq^z - \alpha_Q^z dQ^z + t^z r^z\cdot\delta L^{zi}$$

Similar reasoning establishes the corresponding results for the welfare change to society:

$$(21)\quad dW = \left[\alpha_q^z - \frac{\partial\Gamma^z}{\partial q}\right] dq^z$$

$$- \sum_k \alpha_Q^k dQ^k + \sum_k t^k r^k\cdot\delta L^{ki}$$

$$(22)\quad dW^0 = \sum_{k\neq z}\left[t^k r^k\cdot\delta L^{ki} - \alpha_Q^k dQ^k\right]$$

The behavior of the model is now very much as it was in the case of lump sum taxation; changes in the tax base matter in the same way in both situations. There is one important difference, however. In the lump sum case, new residents automatically brought in some new tax base, whereas now, they only generate new tax base if they increase the demand for taxable land.

VI. The Optimal Dispersion of Economic Activity

How should we evaluate the performance of the models in Sections III–V? So far, this paper has been confined to conditional statements. For example, we can say that if social policy is restricted to choices of public service levels and Tiebout-type lump sum taxes with all other decisions being made according to free-market choice, then the gap between the private and social incentive to expand will be as measured in (10). If communities behave in a Nash way with respect to one another (rather than coordinating activities), this gap in incentives will lead to misallocation of resources relative to what would be second best optimal. The importance of the misallocation clearly depends on the ability of communities to attract residents (through improvement of public services) since the gap is proportional to induced migrations. Although we cannot evaluate the migration functions directly, we can discuss intuitively the nature of the misallocation.

Whenever tax rates exceed (on average) marginal congestion costs, (10) tells us that Nash communities will tend to be too expansionary, engaging in expansionist projects which would not be desirable to the society as a whole. Of course, if tax rates fall short (on average) of marginal congestion costs, myopic communities will tend to be too contractionary. These forces presumably will lead to an adjustment process in which the number and size distribution of communities will change as some communities lose out to more successful and/or aggressive neighbors. Since expansion will tend to increase marginal congestion costs, and vice versa, I suggest that these forces will lead toward an outcome in which tax rates equal marginal congestion costs. If such an outcome were somehow reached, the private and social marginal incentives would be in agreement.

Of course, the situation is quite different if we pass to the model of Section IV in which all taxation in the community must be ad valorem property taxation on a *fixed* tax base. With the world described in that way, communities always have an incentive to *underexpand* (relative to what is *now* second best optimal) as long as there is any congestion at all on public facilities; each community gains nothing from the addition of a new resident (the tax base is not expanded) but must incur the congestion cost. To the extent that communities are successful in restricting entry by controlling the level of public services, we expect an outcome involving relatively small uncongested communities.

The reason that the two analyses just given can differ so markedly in their conclusions is that the concept of the second best optimum is completely different in the two cases. Indeed, the second best optimum in Section IV is quite restrictive relative to that of Section III (or V). With the tax base and size of the community (as determined by the taxed land area) fixed (as in Section IV), there is very little latitude to society in determining the relative size of communities; consequently, the second best optimum is not very interesting from a normative point of view.

The second best optimum of Section V is closer to a first best. There, the tax base of a community is variable, and the division of land use in an area between concentrations of economic activity (communities) and relatively "self-sufficient" use (farms) is an endogenous choice. We can even think of the number of towns as being a variable in this model; a community will disappear if its tax rates and public service levels reach some minimal level, so, as long as we initialize the model with a sufficiently large collection of potential communities, the number of communities is endogenous to the model.

However, no matter how careful we are in generating latitude in the size and composition of communities by refining the models of Sections III and V, the welfare optimum will still be second best, as long as the tax system is constrained to treat equals equally. The first best optimum cannot be decentralized unless equals are treated unequally.[18]

Let us reexamine the congestion tax with this fact in mind. When this tax is consistently charged, there is no marginal social incentive to expand or contract any community, according to the measures developed above. Does this mean that such taxes are optimal? Interestingly, congestion taxes are known to be first best optimal for determining the best size for a nonspatial club (see Buchanan, 1965, who was the first to show this); members should be added to such a club until the congestion cost imposed by a new addition just offsets the cost-spreading benefits derived from new tax payments.

---

[18] Compare Mirrlees.

However, it has been determined recently that this rule is definitely *not* optimal for determining the best size for a *spatial* club of the type studied here (see my 1974 paper). Intuitively, the optimal spatial club trades off the increasing returns aspects of tax spreading (measured by the tax revenue) against the diminishing returns aspects of congestion (measured by the congestion costs) *plus* the diminishing returns effect of increased average transport cost within the town as town size increases (measured by some fraction of the transport costs).

Therefore, at the true first best welfare optimum, the cost of the public goods should exceed congestion costs; indeed, I showed in my 1974 paper that a congestion tax *together with a 100 percent tax on differential land rent* would be optimal. And if communities expand until their congestion costs are as large as the cost of public goods (as they will tend to in the models of Sections III and V), they clearly will be too large relative to what is first best.

There is a good intuitive argument why communities will ignore the increased transport costs in their calculations. Residents (who are the decision-making group) do not pay the "market" costs of expansion (only the congestion costs). Those who move in pay the market costs; either they pay the increased transport costs directly (if they live on the outskirts) or they pay them indirectly by moving into the center and paying a higher rent to a resident (owner) who moves out. Established owners make money on land and will encourage new residents even when it would be better from the point of view of society for a new community to form.

Assuming that communities coordinate their activities and make decisions on the basis of equation (21), there is still a bias from the failure to tax equals unequally, although the direction of the bias is harder to determine. However, we have seen that it is not desirable to tax at marginal congestion costs, even though such a policy would appear to correct for the second best distortion in equation (9). Fortunately, there is no paradox here. It is well known from the general theory of second best that it is not necessarily desirable to eliminate distortions from one part of the economy if there are other distortions elsewhere. Here, the relevant "other" distortion derives from the fact that we have treated equals equally. When an individual moves from a small community to the boundary of a large one, his or her marginal utility of income goes down since part of any additional dollar is wasted on higher rents or added transport. The "extra" transport cost which is a concomitant of larger cities is ignored in the second best welfare measures of Section III.

VII. Conclusions

I have explored a model in which communities compete for residents in a Nash way. I showed that such communities may have a bias toward over- or underexpansion depending on the nature of taxation. Communities restricted to the use of property taxes on a *fixed* tax base will try to be more exclusive than is first best optimal; the prototype for such a community would be a particular suburb within a larger metropolitan area. I would expect such communities to follow restrictive-entry policies to the extent that they are free to act independently.

On the other hand, communities which can employ more general types of taxation, or for which the property tax base is potentially variable, have an incentive to overexpand relative to the first best optimum. The prototype for such a community would be a metropolitan area (to the extent that it is able to coordinate the planning of its subcommunities).

REFERENCES

James Buchanan, "An Economic Theory of Clubs," *Economica,* Feb. 1965, *32,* 1–14.

______, *The Demand and Supply of Public Goods,* Chicago 1968.

______ **and C. Goetz,** "Efficiency Limits of

Fiscal Mobility: An Assessment of the Tiebout Model," *J. Publ. Econ.*, Mar. 1972, *1*, 25–44.

W. E. Diewert, "Applications of Duality Theory," in Michael Intrilligator and David Kendrick, eds., *Frontiers of Quantitative Economics*, Vol. 2, Amsterdam 1976.

F. Flatters, V. Henderson, and P. Mieszkowski, "Public Goods, Efficiency and Regional Fiscal Equalization," *J. Publ. Econ.*, May 1974, *3*, 99–112.

B. Hamilton, "Property Taxes and the Tiebout Hypothesis: Some Empirical Evidence," in Edwin Mills and Wallace Oates, eds., *Fiscal Zoning and Land Use Controls*, Lexington 1974.

J. Mirrlees, "The Optimal Town," *Swedish J. Econ.*, Mar. 1972, *74*, 114–35.

D. Starrett, "Principles of Optimal Location in a Large Homogeneous Area," *J. Econ. Theory*, Dec. 1974, *9*, 418–48.

______, "Measuring Externalities and Second Best Distortions in the Theory of Local Public Goods," *Econometrica*, forthcoming.

C. Tiebout, "A Pure Theory of Local Expenditure," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

B. Weisbrod, "Income Redistribution Effects and Benefit-Cost Analysis," in Samuel Chase, ed., *Problems of Public Expenditure Analysis*, Washington 1972.

F. Westhoff, "Existence of Equilibria in Economics with a Local Public Good," *J. Econ. Theory*, Feb. 1977, *14*, 84–112.

R. Willig, "Consumer's Surplus Without Apology," *Amer. Econ. Rev.*, Sept. 1976, *66*, 589–97.

On the Impossibility of Informationally Efficient Markets

By SANFORD J. GROSSMAN AND JOSEPH E. STIGLITZ*

If competitive equilibrium is defined as a situation in which prices are such that all arbitrage profits are eliminated, is it possible that a competitive economy always be in equilibrium? Clearly not, for then those who arbitrage make no (private) return from their (privately) costly activity. Hence the assumptions that all markets, including that for information, are always in equilibrium and always perfectly arbitraged are inconsistent when arbitrage is costly.

We propose here a model in which there is an equilibrium degree of disequilibrium: prices reflect the information of informed individuals (arbitrageurs) but only partially, so that those who expend resources to obtain information do receive compensation. How informative the price system is depends on the number of individuals who are informed; but the number of individuals who are informed is itself an endogenous variable in the model.

The model is the simplest one in which prices perform a well-articulated role in conveying information from the informed to the uninformed. When informed individuals observe information that the return to a security is going to be high, they bid its price up, and conversely when they observe information that the return is going to be low. Thus the price system makes publicly available the information obtained by informed individuals to the uniformed. In general, however, it does this imperfectly; this is perhaps lucky, for were it to do it perfectly, an equilibrium would not exist.

In the introduction, we shall discuss the general methodology and present some conjectures concerning certain properties of the equilibrium. The remaining analytic sections of the paper are devoted to analyzing in detail an important example of our general model, in which our conjectures concerning the nature of the equilibrium can be shown to be correct. We conclude with a discussion of the implications of our approach and results, with particular emphasis on the relationship of our results to the literature on "efficient capital markets."

I. The Model

Our model can be viewed as an extension of the noisy rational expectations model introduced by Robert Lucas and applied to the study of information flows between traders by Jerry Green (1973); Grossman (1975, 1976, 1978); and Richard Kihlstrom and Leonard Mirman. There are two assets: a safe asset yielding a return R, and a risky asset, the return to which, u, varies randomly from period to period. The variable u consists of two parts,

$$u=\theta+\varepsilon \tag{1}$$

where θ is observable at a cost c, and ε is unobservable.[1] Both θ and ε are random variables. There are two types of individuals, those who observe θ (informed traders), and those who observe only price (uninformed traders). In our simple model, all individuals are, *ex ante*, identical; whether they are informed or uninformed just depends on whether they have spent c to obtain information. Informed traders' demands will depend on θ and the price of the risky asset P. Uninformed traders' demands

---

*University of Pennsylvania and Princeton University, respectively. Research support under National Science Foundation grants SOC76-18771 and SOC77-15980 is gratefully acknowledged. This is a revised version of a paper presented at the Econometric Society meetings, Winter 1975, at Dallas, Texas.

[1] An alternative interpretation is that θ is a "measurement" of u with error. The mathematics of this alternative interpretation differ slightly, but the results are identical.

will depend only on P, but we shall assume that they have rational expectations; they learn the relationship between the distribution of return and the price, and use this in deriving their demand for the risky assets. If x denotes the supply of the risky asset, an equilibrium when a given percentage, λ, of traders are informed, is thus a price function $P_{\lambda}(\theta,x)$ such that, when demands are formulated in the way described, demand equals supply. We assume that uninformed traders do not observe x. Uninformed traders are prevented from learning θ via observations of $P_{\lambda}(\theta,x)$ because they cannot distinguish variations in price due to changes in the informed trader's information from variations in price due to changes in aggregate supply. Clearly, $P_{\lambda}(\theta,x)$ reveals some of the informed trader's information to the uninformed traders.

We can calculate the expected utility of the informed and the expected utility of the uninformed. If the former is greater than the latter (taking account of the cost of information), some individuals switch from being uninformed to being informed (and conversely). An overall equilibrium requires the two to have the same expected utility. As more individuals become informed, the expected utility of the informed falls relative to the uninformed for two reasons:

(a) The price system becomes more informative because variations in θ have a greater effect on aggregate demand and thus on price when more traders observe θ. Thus, more of the information of the informed is available to the uninformed. Moreover, the informed gain more from trade with the uninformed than do the uninformed. The informed, on average, buy securities when they are "underpriced" and sell them when they are "overpriced" (relative to what they would have been if information were equalized).[2] As the price system becomes more informative, the difference in their information—and hence the magnitude by which the informed can gain relative to the uninformed—is reduced.

(b) Even if the above effect did not occur, the increase in the ratio of informed to uninformed means that the relative gains of the informed, on a per capita basis, in trading with the uninformed will be smaller.

We summarize the above characterization of the equilibrium of the economy in the following two conjectures:

Conjecture 1: The more individuals who are informed, the more informative is the price system.

Conjecture 2: The more individuals who are informed, the lower the ratio of expected utility of the informed to the uninformed.

(Conjecture 1 obviously requires a definition of "more informative"; this is given in the next section and in fn. 7.)

The equilibrium number of informed and uninformed individuals in the economy will depend on a number of critical parameters: the cost of information, how informative the price system is (how much noise there is to interfere with the information conveyed by the price system), and how informative the information obtained by an informed individual is.

Conjecture 3: The higher the cost of information, the smaller will be the equilibrium percentage of individuals who are informed.

Conjecture 4: If the quality of the informed trader's information increases, the more their demands will vary with their information and thus the more prices will vary with θ. Hence, the price system becomes more informative. The equilibrium proportion of informed to uninformed may be either increased or decreased, because even though the value of being informed has increased due to the increased quality of θ, the value of being uninformed has also increased because the price system becomes more informative.

Conjecture 5: The greater the magnitude of noise, the less informative will the price system be, and hence the lower the expected utility of uninformed individuals. Hence, in equilibrium the greater the magnitude of noise, the larger the proportion of informed individuals.

---

[2] The framework described herein does not explicitly model the effect of variations in supply, i.e., x on commodity storage. The effect of futures markets and storage capabilities on the informativeness of the price system was studied by Grossman (1975, 1977).

Conjecture 6: In the limit, when there is no noise, prices convey all information, and there is no incentive to purchase information. Hence, the only possible equilibrium is one with no information. But if everyone is uninformed, it clearly pays some individual to become informed.[3] Thus, there does not exist a competitive equilibrium.[4]

Trade among individuals occurs either because tastes (risk aversions) differ, endowments differ, or beliefs differ. This paper focuses on the last of these three. An interesting feature of the equilibrium is that beliefs may be precisely identical in either one of two situations: when all individuals are informed or when all individuals are uninformed. This gives rise to:

Conjecture 7: That, other things being equal, markets will be thinner under those conditions in which the percentage of individuals who are informed (λ) is either near zero or near unity. For example, markets will be thin when there is very little noise in the system (so λ is near zero), or when costs of information are very low (so λ is near unity).

In the last few paragraphs, we have provided a number of conjectures describing the nature of the equilibrium when prices convey information. Unfortunately, we have not been able to obtain a general proof of any of these propositions. What we have been able to do is to analyze in detail an interesting example, entailing constant absolute risk-aversion utility functions and normally distributed random variables. In this example, the equilibrium price distribution can actually be calculated, and all of the conjectures provided above can be verified. The next sections are devoted to solving for the equilibrium in this particular example.[5]

II. Constant Absolute Risk-Aversion Model

A. *The Securities*

The ith trader is assumed to be endowed with stocks of two types of securities: $\overline{M}_i$, the riskless asset, and $\overline{X}_i$, a risky asset. Let P be the current price of risky assets and set the price of risk free assets equal to unity. The ith trader's budget constraint is

$$PX_i + M_i = W_{0i} \equiv \overline{M}_i + P\overline{X}_i \tag{2}$$

Each unit of the risk free asset pays R "dollars" at the end of the period, while each unit of the risky asset pays u dollars. If at the end of the period, the ith trader holds a portfolio (M_i, X_i), his wealth will be

$$W_{1i} = RM_i + uX_i \tag{3}$$

B. *Individual's Utility Maximization*

Each individual has a utility function $V_i(W_{1i})$. For simplicity, we assume all individuals have the same utility function and so drop the subscripts i. Moreover, we assume the utility function is exponential, i.e.,

$$V(W_{1i}) = -e^{-aW_{1i}}, \quad a>0$$

where a is the coefficient of absolute risk aversion. Each trader desires to maximize expected utility, using whatever information is available to him, and to decide on what information to acquire on the basis of the consequences to his expected utility.

Assume that in equation (1) θ and ε have a multivariate normal distribution, with

$$E\varepsilon = 0 \tag{4}$$

$$E\theta\varepsilon = 0 \tag{5}$$

$$Var(u^*|\theta) = Var\,\varepsilon^* \equiv \sigma_\varepsilon^2 > 0 \tag{6}$$

---

[3] That is, with no one informed, an individual can only get information by paying c dollars, since no information is revealed by the price system. By paying c dollars an individual will be able to predict better than the market when it is optimal to hold the risky asset as opposed to the risk-free asset. Thus his expected utility will be higher than an uninformed person gross of information costs. Thus for c sufficiently low all uninformed people will desire to be informed.

[4] See Grossman (1975, 1977) for a formal example of this phenomenon in futures markets. See Stiglitz (1971, 1974) for a general discussion of information and the possibility of nonexistence of equilibrium in capital markets.

[5] The informational equilibria discussed here may not, in general, exist. See Green (1977). Of course, for the utility function we choose equilibrium does exist.

since θ and ε are uncorrelated. Throughout this paper we will put a * above a symbol to emphasize that it is a random variable. Since W_{1i} is a linear function of ε, for a given portfolio allocation, and a linear function of a normally distributed random variable is normally distributed, it follows that W_{1i} is normal conditional on θ. Then, using (2) and (3) the expected utility of the *informed* trader with information θ can be written

$$(7)\quad E(V(W^*_{1i})|\theta)= -\exp\left(-a\left\{E[W^*_{1i}|\theta]-\frac{a}{2}Var[W^*_{1i}|\theta]\right\}\right)$$

$$=-\exp\left(-a\left[RW_{0i}+X_I\{E(u^*|\theta)-RP\} -\frac{a}{2}X_I^2\,Var(u^*|\theta)\right]\right)$$

$$=-\exp\left(-a\left[RW_{0i}+X_I(\theta-RP) -\frac{a}{2}X_I^2\sigma_\varepsilon^2\right]\right)$$

where X_I is an informed individual's demand for the risky security. Maximizing (7) with respect to X_I yields a demand function for risky assets:

$$(8)\qquad X_I(P,\theta)=\frac{\theta-RP}{a\sigma_\varepsilon^2}$$

The right-hand side of (8) shows the familiar result that with constant absolute risk aversion, a trader's demand does not depend on wealth; hence the subscript i is not on the left-hand side of (8).

We now derive the demand function for the uninformed. Let us assume the only source of "noise" is the per capita supply of the risky security x.

Let $P^*(\cdot)$ be some particular price function of (θ,x) such that u^* and P^* are jointly normally distributed. (We will prove that this exists below.)

Then, we can write for the uninformed individual

$$(7')\quad E(V(W^*_{1i})|P^*)=-\exp\left[-a\left\{E[W^*_{1i}|P^*] -\frac{a}{2}Var[W^*_{1i}|P^*]\right\}\right]$$

$$=-\exp\left[-a\left\{RW_{0i}+X_U(E[u^*|P^*]-RP) -\frac{a}{2}X_U^2\,Var[u^*|P^*]\right\}\right]$$

The demands of the uninformed will thus be a function of the price function P^* and the actual price P.

$$(8')\quad X_U(P;P^*) =\frac{E[u^*|P^*(\theta,x)=P]-RP}{a\,Var[u^*|P^*(\theta,x)=P]}$$

C. *Equilibrium Price Distribution*

If λ is some particular fraction of traders who decide to become informed, then define an equilibrium price system as a function of $(\theta,x), P_\lambda(\theta,x)$, such that for all (θ,x) per capita demands for the risky assets equal supplies:

$$(9)\qquad \lambda X_I(P_\lambda(\theta,x),\theta) +(1-\lambda)X_U(P_\lambda(\theta,x);P^*_\lambda)=x$$

The function $P_\lambda(\theta,x)$ is a statistical equilibrium in the following sense. If over time uninformed traders observe many realizations of (u^*,P^*_λ), then they learn the joint distribution of (u^*,P^*_λ). After all learning about the joint distribution of (u^*,P^*_λ) ceases, all traders will make allocations and form expectations such that this joint distribution persists over time. This follows from (8), (8′), and (9), where the market-clearing price that comes about is the one which takes into account the fact that uninformed traders have learned that it contains information.

We shall now prove that there exists an equilibrium price distribution such that P^* and u^* are jointly normal. Moreover, we shall be able to characterize the price distribution. We define

$$(10a) \quad w_\lambda(\theta,x) = \theta - \frac{a\sigma_\varepsilon^2}{\lambda}(x - Ex^*)$$

for $\lambda > 0$, and define $w_0(\theta,x)$ as the number:

$$(10b) \quad w_0(\theta,x) = x \qquad \text{for all } (\theta,x)$$

where w_λ is just the random variable θ, plus noise.[6] The magnitude of the noise is inversely proportional to the proportion of informed traders, but is proportional to the variance of ε. We shall prove that the equilibrium price is just a linear function of w_λ. Thus, if $\lambda > 0$, the price system conveys information about θ, but it does so imperfectly.

D. *Existence of Equilibrium and a Characterization Theorem*

THEOREM 1: *If* $(\theta^*,\varepsilon^*,x^*)$ *has a nondegenerate joint normal distribution such that* θ^*, ε^*, *and* x^* *are mutually independent, then there exists a solution to* (9) *which has the form* $P_\lambda(\theta,x) = \alpha_1 + \alpha_2 w_\lambda(\theta,x)$, *where* α_1 *and* α_2 *are real numbers which may depend on* λ, *such that* $\alpha_2 > 0$. (*If* $\lambda = 0$, *the price contains no information about* θ.) *The exact form of* $P_\lambda(\theta,x)$ *is given in equation* (A10) *in Appendix B. The proof of this theorem is also in Appendix B.*

The importance of Theorem 1 rests in the simple characterization of the information in the equilibrium price system: P_λ^* is informationally equivalent to w_λ^*. From (10) w_λ^* is a "mean-preserving spread" of θ; i.e., $E[w_\lambda^*|\theta] = \theta$ and

$$(11) \qquad Var[w_\lambda^*|\theta] = \frac{a^2\sigma_\varepsilon^4}{\lambda^2} Var\,x^*$$

For each replication of the economy, θ is the information that uninformed traders would like to know. But the noise x^* prevents w_λ^* from revealing θ. How well-informed uninformed traders can become from observing P_λ^* (equivalently w_λ^*) is measured by $Var[w_\lambda^*|\theta]$. When $Var[w_\lambda^*|\theta]$ is zero, w_λ^* and θ are perfectly correlated. Hence when uninformed firms observe w_λ^*, this is equivalent to observing θ. On the other hand, when $Var[w_\lambda^*|\theta]$ is very large, there are "many" realizations of w_λ^* that are associated with a given θ. In this case the observation of a particular w_λ^* tells very little about the actual θ which generated it.[7]

From equation (11) it is clear that large noise (high $Var\,x^*$) leads to an imprecise price system. The other factor which determines the precision of the price system $(a^2\sigma_\varepsilon^4/\lambda^2)$ is more subtle. When a is small (the individual is not very risk averse) or σ_ε^2 is small (the information is very precise), an informed trader will have a demand for risky assets which is very responsive to changes in θ. Further, the larger λ is, the more responsive is the total demand of informed traders. Thus small $(a^2\sigma_\varepsilon^4/\lambda^2)$ means that the aggregate demand of informed traders is very responsive to θ. For a fixed amount of noise (i.e., fixed $Var\,x^*$) the larger are the movements in aggregate demand which are due to movements in θ, the more will price movements be due to movements in θ. That is, x^* becomes less important relative to θ in determining price movements. Therefore, for small $(a^2\sigma_\varepsilon^4/\lambda^2)$ uninformed traders are able to confidently know that price is, for example, unusually high due to θ being high. In this way information from informed traders is transferred to uninformed traders.

---

[6] If $y' = y + Z$, and $E[Z|y] = 0$, then y' is just y plus noise.

[7] Formally, w_λ^* is an experiment in the sense of Blackwell which gives information about θ. It is easy to show that, *ceteris paribus*, the smaller $Var(w_\lambda^*|\theta)$ the more "informative" (or sufficient) in the sense of Blackwell, is the experiment; see Grossman, Kihlstrom, and Mirman (p. 539).

E. *Equilibrium in the Information Market*

What we have characterized so far is the equilibrium price distribution for given λ. We now define an *overall* equilibrium to be a pair (λ, P_λ^*) such that the expected utility of the informed is equal to that of the uninformed if $0<\lambda<1$; $\lambda=0$ if the expected utility of the informed is less than that of the uninformed at P_0^*; $\lambda=1$ if the expected utility of the informed is greater than the uninformed at P_1^*. Let

$$\text{(12a)}\quad W_{Ii}^\lambda \equiv R(W_{0i}-c) + [u - RP_\lambda(\theta,x)]X_I(P_\lambda(\theta,x),\theta)$$

$$\text{(12b)}\quad W_{U1}^\lambda \equiv RW_{0i} + [u - RP_\lambda(\theta,x)]X_U(P_\lambda(\theta,x);P_\lambda^*)$$

where c is the cost of observing a realization of θ^*. Equation (12a) gives the end of period wealth of a trader if he decides to become informed, while (12b) gives his wealth if he decides to be uninformed. Note that end of period wealth is random due to the randomness of W_{0i}, u, θ, and x.

In evaluating the expected utility of W_{Ii}^λ, we do not assume that a trader knows which realization of θ^* he gets to observe if he pays c dollars. A trader pays c dollars and then gets to observe some realization of θ^*. The overall expected utility of W_{Ii}^λ averages over all possible θ^*, ε^*, x^*, and W_{0i}. The variable W_{0i} is random for two reasons. First from (2) it depends on $P_\lambda(\theta,x)$, which is random as (θ,x) is random. Secondly, in what follows we will assume that $\bar{X}_i$ is random.

We will show below that $EV(W_{Ii}^\lambda)/EV(W_{Ui}^\lambda)$ is independent of i, but is a function of λ, a, c, and σ_ε^2. More precisely, in Appendix B we prove

THEOREM 2: *Under the assumptions of Theorem* 1, *and if* $\bar{X}_i$ *is independent of* (u^*,θ^*,x^*) *then*

$$\text{(13)}\quad \frac{EV(W_{Ii}^\lambda)}{EV(W_{Ui}^\lambda)} = e^{ac}\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}}$$

F. *Existence of Overall Equilibrium*

Theorem 2 is useful, both in proving the uniqueness of overall equilibrium and in analyzing comparative statics. Overall equilibrium, it will be recalled, requires that for $0<\lambda<1$, $EV(W_{Ii}^\lambda)/EV(W_{Ui}^\lambda)=1$. But from (13)

$$\text{(14)}\quad \frac{EV(W_{Ii}^\lambda)}{EV(W_{Ui}^\lambda)} = e^{ac}\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}} \equiv \gamma(\lambda)$$

Hence overall equilibrium simply requires, for $0<\lambda<1$,

$$\text{(15)}\qquad \gamma(\lambda)=1$$

More precisely, we now prove

THEOREM 3: *If* $0\leqslant\lambda\leqslant 1$, $\gamma(\lambda)=1$, *and* P_λ^* *is given by* (A10) *in Appendix B, then* (λ,P_λ^*) *is an overall equilibrium. If* $\gamma(1)<1$, *then* $(1,P_1^*)$ *is an overall equilibrium. If* $\gamma(0)>1$, *then* $(0,P_0^*)$ *is an overall equilibrium. For all price equilibria* P_λ *which are monotone functions of* w_λ, *there exists a unique overall equilibrium* (λ,P_λ^*).

PROOF:

The first three sentences follow immediately from the definition of overall equilibrium given above equation (12), and Theorems 1 and 2. Uniqueness follows from the monotonicity of $\gamma(\cdot)$ which follows from (A11) and (14). The last two sentences in the statement of the theorem follow immediately.

In the process of proving Theorem 3, we have noted

COROLLARY 1: $\gamma(\lambda)$ is a strictly monotone increasing function of λ.

This looks paradoxical; we expect the ratio of informed to uninformed expected utility to be a decreasing function of λ. But, *we have defined utility as negative*. Therefore as λ rises, the expected utility of informed traders does go down relative to uninformed traders.

Note that the function $\gamma(0)=e^{ac}(Var(u^*|\theta)/Var\, u^*)^{1/2}$. Figure 1 illustrates the determination of the equilibrium λ. The figure assumes that $\gamma(0)<1<\gamma(1)$.

G. *Characterization of Equilibrium*

We wish to provide some further characterization of the equilibrium. Let us define

$$m=\left(\frac{a\sigma_\varepsilon^2}{\lambda}\right)^2\frac{\sigma_x^2}{\sigma_\theta^2} \tag{16a}$$

$$n=\frac{\sigma_\theta^2}{\sigma_\varepsilon^2} \tag{16b}$$

Note that m is inversely related to the informativeness of the price system since the squared correlation coefficient between P_λ^* and θ^*, ρ_θ^2 is given by

$$\rho_\theta^2=\frac{1}{1+m} \tag{17}$$

Similarly, n is directly related to the quality of the informed trader's information because $n/(1+n)$ is the squared correlation coefficient between θ^* and u^*.

Equations (14) and (15) show that the cost of information c, determines the equilibrium ratio of information quality between informed and uninformed traders $(Var(u^*|\theta))/Var(u^*|w_\lambda)$. From (1), (A11) of Appendix A, and (16), this can be written as

$$\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}=\frac{1+m}{1+m+nm}=\left(1+\frac{nm}{1+m}\right)^{-1} \tag{18}$$

Substituting (18) into (14) and using (15) we obtain, for $0<\lambda<1$, in equilibrium

$$m=\frac{e^{2ac}-1}{1+n-e^{2ac}} \tag{19a}$$

or

$$1-\rho_\theta^2=\frac{e^{2ac}-1}{n} \tag{19b}$$

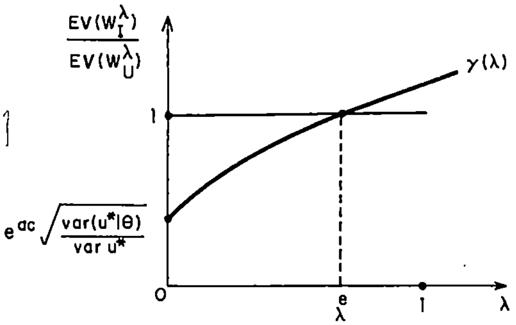


FIGURE 1

Note that (19) holds for $\gamma(0)<1<\gamma(1)$, since these conditions insure that the equilibrium λ is between zero and one. Equation (19b) shows that the equilibrium informativeness of the price system is determined completely by the cost of information c, the quality of the informed trader's information n, and the degree of risk aversion a.

H. *Comparative Statics*

From equation (19b), we immediately obtain some basic comparative statics results:

1) An increase in the quality of information (n) increases the informativeness of the price system.

2) A decrease in the cost of information increases the informativeness of the price system.

3) A decrease in risk aversion leads informed individuals to take larger positions, and this increases the informativeness of the price system.

Further, all other changes in parameters, such that n, a, and c remain constant, do not change the equilibrium degree of informativeness of the price system; other changes lead only to particular changes in λ of a magnitude to exactly offset them. For example:

4) An increase in noise (σ_x^2) increases the proportion of informed traders. At any given λ, an increase in noise reduces the informativeness of the price system; but it increases the returns to information and leads more individuals to become informed; the remarkable result obtained above establishes that *the two effects exactly offset each*

other so that the equilibrium informativeness of the price system is unchanged. This can be illustrated diagrammatically if we note from (16a) that for a given λ, an increase in σ_x^2 raises m which from (18) lowers $(Var(u^*|\theta))/Var(u^*|w_\lambda)$. Thus from (14) a rise in σ_x^2 leads to a vertical downward shift of the $\gamma(\lambda)$ curve in Figure 1, and thus a higher value of λ^e.

5) Similarly an increase in σ_ε^2 for a constant n (equivalent to an increase in the variance of u since n is constant) leads to an increased proportion of individuals becoming informed—and indeed again just enough to offset the increased variance, so that the degree of *informativeness* of the price system remains unchanged. This can also be seen from Figure 1 if (16) is used to note that an increase in σ_ε^2 with n held constant by raising σ_θ^2 leads to an increase in m for a given λ. From (18) and (14) this leads to a vertical downward shift of the $\gamma(\lambda)$ curve and thus a higher value of λ^e.

6) It is more difficult to determine what happens if, say σ_θ^2 increases, keeping σ_u^2 constant (implying a fall in σ_ε^2), that is, *the information obtained is more informative.* This leads to an increase in n, which from (19b) implies that the equilibrium informativeness of the price system rises. From (16) it is clear that m and nm both fall when σ_θ^2 rises (keeping $\sigma_u^2 = \sigma_\theta^2 + \sigma_\varepsilon^2$ constant). This implies that the $\gamma(\lambda)$ curve may shift up or down depending on the precise values of c, a, and n.[8] This ambiguity arises because an improvement in the precision of informed traders' information, with the cost of the information fixed, increases the benefit of being informed. However, some of the improved information is transmitted, via a more informative price system, to the uninformed; this increases the benefits of being uninformed. If n is small, both the price system m is not very informative *and* the marginal value of information to informed traders is high. Thus the *relative* benefits of being informed rises when n rises; implying that the equilibrium λ rises. Conversely when n is large the price system is very informative and the marginal value of information is low to informed traders so the relative benefits of being uninformed rises.

7) From (14) it is clear that an increase in the cost of information c shifts the $\gamma(\lambda)$ curve up and thus decreases the percentage of informed traders.

The above results are summarized in the following theorem.

THEOREM 4: *For equilibrium* λ *such that* $0<\lambda<1$:

A. *The equilibrium informativeness of the price system,* ρ_θ^2, *rises if* n *rises,* c *falls, or* a *falls.*

B. *The equilibrium informativeness of the price system is unchanged if* σ_x^2 *changes, or if* σ_u^2 *changes with* n *fixed.*

C. *The equilibrium percentage of informed traders will rise if* σ_x^2 *rises,* σ_u^2 *rises for a fixed* n, *or* c *falls.*

D. *If* $\bar{n}$ *satisfies* $(e^{2ac}-1)/(\bar{n}-(e^{2ac}-1)) = \bar{n}/(\bar{n}+1)$, *then* $n \overset{(<)}{>} \bar{n}$ *implies that* λ *falls (rises) due to an increase in* n.

PROOF:

Parts A–C are proved in the above remarks. Part D is proved in footnote 8.

I. *Price Cannot Fully Reflect Costly Information*

We now consider certain limiting cases, for $\gamma(0) \leqslant 1 \leqslant \gamma(1)$, and show that equilibrium does not exist if $c>0$ and price is fully informative.

1) As the cost of information goes to zero, the price system becomes more infor-

---

[8] From (14) and (18) it is clear that λ rises if and only if $Var(u^*|\theta) + Var(u^*|w_\lambda)$ falls due to the rise in σ_θ^2 for a given λ. This occurs if and only if $nm/(1+m)$ rises. Using (16) to differentiate $nm/(1+m)$ with respect to σ_ε^2 subject to the constraint that $d\sigma_u^2 = 0$ (i.e., $d\sigma_\theta^2 = -d\sigma_\varepsilon^2$), we find that the sign of

$$\frac{d}{d\sigma_\theta^2}\left(\frac{nm}{1+m}\right) = sgn\left[m\left(\frac{n+1}{n}\right)-1\right]$$

$$= sgn\left[\left(\frac{\gamma}{n-\gamma}\right)\left(\frac{n+1}{n}\right)-1\right]$$

where $\gamma \equiv e^{2ac}-1$ and the last equality follows from equation (19a). Thus for n very large the derivative is negative so that λ falls due to an increase in the precision of the informed trader's information. Similarly if n is sufficiently small, the derivative is positive and thus λ rises.

mative, but at a positive value of c, say $\hat{c}$, all traders are informed. From (14) and (15) $\hat{c}$ satisfies

$$e^{a\hat{c}}\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_1)}}=1$$

2) From (19a) as the precision of the informed trader's information n goes to infinity, i.e., $\sigma_\varepsilon^2 \to 0$ and $\sigma_\theta^2 \to \sigma_u^2$, σ_u^2 held fixed, the price system becomes perfectly informative. Moreover the percentage of informed traders goes to zero! This can be seen from (18) and (15). That is, as $\sigma_\varepsilon^2 \to 0$, $nm/(1+m)$ must stay constant for equilibrium to be maintained. But from (19b) and (17), m falls as σ_ε^2 goes to zero. Therefore nm must fall, but nm must not go to zero or else $nm/(1+m)$ would not be constant. From (16) $nm=(a/\lambda)^2\sigma_\varepsilon^2\sigma_x^2$, and thus λ must go to zero to prevent nm from going to zero as $\sigma_\varepsilon^2 \to 0$.

3) From (16a) and (19a) it is clear that as noise σ_x^2 goes to zero, the percentage of informed traders goes to zero. Further, since (19a) implies that m does not change as σ_x^2 changes, the informativeness of the price system is unchanged as $\sigma_x^2 \to 0$.

Assume that c is small enough so that it is worthwhile for a trader to become informed when no other trader is informed. Then if $\sigma_x^2=0$ or $\sigma_\varepsilon^2=0$, there exists no competitive equilibrium. To see this, note that equilibrium requires either that the ratio of expected utility of the informed to the uninformed be equal to unity, or that if the ratio is larger than unity, no one be informed. We shall show that when no one is informed, it is less than unity so that $\lambda=0$ cannot be an equilibrium; but when $\lambda>0$, it is greater than unity. That is, if $\sigma_x^2=0$ or $\sigma_\varepsilon^2=0$, the ratio of expected utilities is not a continuous function of λ at $\lambda=0$.

This follows immediately from observing that at $\lambda=0$, $Var(u^*|w_0)=Var\,u^*$, and thus by (14)

$$\frac{EV(W_{Ii}^0)}{EV(W_{Ui}^0)}=e^{ac}\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}} \tag{20}$$

$$=e^{ac}\sqrt{\frac{1}{1+n}}$$

while if $\lambda>0$, by (18)

$$\frac{EV(W_{Ii}^\lambda)}{EV(W_{Ui}^\lambda)}=e^{ac}\sqrt{\frac{1}{1+n\dfrac{m}{m+1}}}$$

But if $\sigma_x^2=0$ or $\sigma_\varepsilon^2=0$, then $m=0$, $nm=0$ for $\lambda>0$, and hence

$$\lim_{\lambda\to 0}\frac{EV(W_{Ii}^\lambda)}{EV(W_{Ui}^\lambda)}=e^{ac} \tag{21}$$

It immediately follows that

THEOREM 5: (a) *If there is no noise* ($\sigma_x^2=0$), *an overall equilibrium does not exist if* (*and only if*) $e^{ac}<\sqrt{1+n}$. (b) *If information is perfect* ($\sigma_\varepsilon^2=0, n=\infty$), *there never exists an equilibrium.*

PROOF:

(a) If $e^{ac}<\sqrt{1+n}$, then by (20) and (21), $\gamma(\lambda)$ is discontinuous at $\lambda=0$; $\lambda=0$ is not an equilibrium since by (20) $\gamma(0)<1$; $\lambda>0$ is not an equilibrium since by (21) $\gamma(\lambda)>1$.

(b) If $\sigma_\varepsilon^2=0$ and $\sigma_\theta^2=\sigma_u^2$ so that information is perfect, then for $\lambda>0$, $nm=0$ by (16) and hence $\gamma(\lambda)>1$ by (21). From (20) $\gamma(0)=0<1$.

If there is no noise and some traders become informed, then *all* their information is transmitted to the uninformed by the price system. Hence each informed trader acting as a price taker thinks the informativeness of the price system will be unchanged if he becomes uninformed, so $\lambda>0$ is not an equilibrium. On the other hand, if no traders are informed, then each uninformed trader learns nothing from the price system, and thus he has a desire to become informed (if $e^{ac}<(1+n)^{1/2}$). Similarly if the informed traders get perfect information, then their demands are very sensitive to their information, so that the market-clearing price becomes very sensitive to their information and thus reveals θ to the uninformed. Hence all traders desire to be uninformed. But if all traders are uninformed, each trader can eliminate the risk of his portfolio by the purchase of information, so each trader desires to be informed.

In the next section we show that the nonexistence of competitive equilibrium can be thought of as the breakdown of competitive markets due to lack of trade. That is, we will show that as σ_x^2 gets very small, trade goes to zero and markets serve no function. Thus competitive markets close for lack of trade "before" equilibrium ceases to exist at $\sigma_x^2=0$.

III. On the Thinness of Speculative Markets

In general, trade takes place because traders differ in endowments, preferences, or beliefs. Grossman (1975, 1977, 1978) has argued that differences in preferences are not a major factor in explaining the magnitude of trade in speculative markets. For this reason the model in Section II gave all traders the same risk preferences (note that none of the results in Section II are affected by letting traders have different coefficients of absolute risk aversion). In this section we assume that trade requires differences in endowments or beliefs and dispense with differences in risk preference as an explanatory variable.[9]

There is clearly some fixed cost in operating a competitive market. If traders have to bear this cost, then trade in the market must be beneficial. Suppose traders have the same endowments and beliefs. Competitive equilibrium will leave them with allocations which are identical with their initial endowments. Hence, if it is costly to enter such a competitive market, no trader would ever enter. We will show below that in an important class of situations, there is continuity in the amount of net trade. That is, when initial endowments are the same and peoples' beliefs differ *slightly*, then the competitive equilibrium allocation that an individual gets will be only *slightly* different from his initial endowment. Hence, there will only be a slight benefit to entering the competitive market. This could, for sufficiently high operating costs, be outweighed by the cost of entering the market.

The amount of trade occurring at any date is a random variable; a function of θ and x. It is easy to show that it is a normally distributed random variable. Since one of the primary determinants of the size of markets is differences in beliefs, one might have conjectured that markets will be thin, in some sense, if almost all traders are either informed or uninformed. This is not, however, obvious, since the amount of trade by any single trader may be a function of λ as well, and a few active traders can do the job of many small traders. In our model, there is a sense, however, in which our conjecture is correct.

We first calculate the magnitude of trades as a function of the exogenous parameters, θ and x. Let $h\equiv\sigma_\epsilon^2$, $\bar{x}=Ex^*$, and $\bar{\theta}\equiv E\theta^*$. (The actual trades will depend on the distribution of random endowments across all of the traders, but these we shall net out.) Per capita net trade is [10]

$$\text{(22)}\quad X_I-x=(1-\lambda)\left[\left(nm+\frac{ah}{\lambda}\right)(x-\bar{x})\right.$$
$$\left.+\left[(m+1)n-1\right](\theta-\bar{\theta})+\bar{x}nm\right]$$
$$\div\left[1+m+\lambda nm\right]$$

---

[9] In the model described in Section II it was assumed that an individual's endowment $\bar{X}_i$ is independent of the market's per capita endowment x^*. This was done primarily so there would not be useful information in an individual's endowment about the total market endowment. Such information would be useful in equilibrium because an individual observes $P_\lambda(\theta,x)$. If due to observing $\bar{X}_i$, he knows something about x, then by observing $P_\lambda(\theta,x)$, $\bar{X}_i$ is valuable in making inferences about θ. To take this into account is possible, but would add undue complication to a model already overburdened with computations.

[10] Calculation of distribution of net trades

$$\frac{\lambda}{ah}(\theta-RP_\lambda)$$
$$+\frac{(1\text{-}\lambda)\left[\left(\bar{\theta}-RP_\lambda\right)(1+m)n+\theta-\bar{\theta}-\frac{ah}{\lambda}(x-\bar{x})\right]}{ah(1+m+nm)n}=x$$
$$\text{or}\quad\frac{(\theta-RP_\lambda)}{ah}\left(\lambda+\frac{(1-\lambda)(1+m)}{1+m+nm}\right)$$
$$=\left(\frac{\theta-RP_\lambda}{ah}\right)\left(\frac{1+m+\lambda nm}{1+m+nm}\right)$$
$$=x+\frac{(1-\lambda)\left([(m+1)n-1](\theta-\bar{\theta})+\frac{ah}{\lambda}(x-\bar{x})\right)}{ah(1+m+\lambda nm)n}$$

Thus, the mean of total informed trade is

$$E\lambda(X_I - x) = \frac{(1-\lambda)\lambda m\bar{x}}{1+m+\lambda nm} \tag{23}$$

and its variance is

$$\sigma_\theta^2(1-\lambda)^2\lambda^2\left[[(m+1)n-1]^2 + \left(nm+\frac{a\sigma_\varepsilon^2}{\lambda}\right)^2\frac{\sigma_x^2}{\sigma_\theta^2}\right] \div (1+m+\lambda nm)^2 n^2 \tag{24}$$

In the last section we considered limiting values of the exogenous variables with the property that $\lambda \to 0$. The following theorem will show that the mean and variance of trade go to zero as $\lambda \to 0$. That is, the distribution of $\lambda(X_I - x)$ becomes degenerate at zero as $\lambda \to 0$. This is not trivial because as $\lambda \to 0$ due to $n \to \infty$ (very precise information), the informed trader's demand $X_I(P,\theta)$ goes to infinity at most prices because the risky asset becomes riskless with perfect information.

THEOREM 6: (a) *For sufficiently large or small c, the mean and variance of trade is zero.* (b) *As the precision of informed traders' information n goes to infinity, the mean and variance of trade go to zero.*

PROOF:

(a) From remark 1) in Section II, Part I, $\lambda = 1$ if $c \leqslant \hat{c}$, which from (23) and (24) implies trade is degenerate at zero. From (14), for c sufficiently large, say c^0, $\gamma(0)=1$, so the equilibrium $\lambda = 0$. As c goes to c^0 from below $\lambda \to 0$, and from (14), (15), and (18) $\lim_{c\uparrow c^0}(1+nm/(1+m))^{-1/2} = e^{-ac^0}$. Hence $\lim_{c\uparrow c^0}(nm/1+m)$ is a finite positive number. Thus from (22) mean trade goes to zero as $c\uparrow c^0$. If the numerator and the denominator of (24) are divided by $(1+m)^2$, then again using the fact that $m/1+m$ has a finite limit gives the result that as $c\uparrow c^0$, $\lambda \to 0$, and variance of trade goes to zero.

(b) By (14), (15), and (18), $nm/(1+m)$ is constant as $n \to \infty$. Further, from remark 2) of Section II, Part I, $\lambda \to 0$ as $n \to \infty$. Hence from (23) and (24), the mean and variance of trade go to zero.

(c) From remark 3) in Section II, Part I, m is constant and λ goes to zero as $\sigma_x^2 \to 0$. Therefore mean trade goes to zero. In (24), note that $(nm + a\sigma_\varepsilon^2/\lambda)^2\sigma_x^2/\sigma_\theta^2 = (nm\sigma_x/\sigma_\theta + (m)^{1/2})^2$ by (16a). Hence the variance of trade goes to zero as $\sigma_x^2 \to 0$.

Note further that $\lambda(X_I - x) + (1-\lambda)(X_U - x) = 0$ implies that no trade will take place as $\lambda \to 1$. Thus, the result that competitive equilibrium is incompatible with informationally efficient markets should be interpreted as meaning that speculative markets where prices reveal a lot of information will be very thin because it will be composed of individuals with very similar beliefs.

IV. On the Possibility of Perfect Markets

In Section II we showed that the price system reveals the signal w_λ^* to traders, where

$$w_\lambda \equiv \theta - \frac{a\sigma_\varepsilon^2}{\lambda}(x - Ex^*)$$

Thus, for given information of informed traders θ, the price system reveals a noisy version of θ. The noise is $(a\sigma_\varepsilon^2/\lambda)(x-Ex^*)$. Uninformed traders learn θ to within a random variable with mean zero and variance $(a\sigma_\varepsilon^2/\lambda)^2\,Var\,x^*$, where σ_ε^2 is the precision of informed traders' information, $Var\,x^*$ is the amount of endowment uncertainty, λ the fraction of informed traders, and a is the degree of absolute risk aversion. Thus, in general the price system does not reveal all

---

or $$X_I = \frac{1+m+nm}{1+m+\lambda nm} \times\left[x + \frac{(1-\lambda)\left([(m+1)-1](\theta-\bar{\theta}) + \frac{ah}{\lambda}(x-\bar{x})\right)}{ah(1+m+nm)n}\right]$$

$$X_I - x = \frac{(1-\lambda)\left[\left(nm+\frac{ah}{\lambda}\right)(x-\bar{x}) + [(m+1)-1](\theta-\bar{\theta}) + \bar{x}nm\right.}{(1+m+\lambda nm)n}$$

the information about "the true value" of the risky asset. (θ is the true value of the risky asset in that it reflects the best available information about the asset's worth.)

The only way informed traders can earn a return on their activity of information gathering, is if they can use their information to take positions in the market which are "better" than the positions of uninformed traders. "Efficient Markets" theorists have claimed that "at any time prices fully reflect all available information" (see Eugene Fama, p. 383). If this were so then informed traders could not earn a return on their information.

We showed that when the efficient markets hypothesis is true and information is costly, competitive markets break down. This is because when $\sigma_\epsilon^2=0$ or $Var\, x^*=0$, w_λ, and thus price, does reflect all the information. When this happens, each informed trader, because he is in a competitive market, feels that he could stop paying for information and do as well as a trader who pays nothing for information. But all informed traders feel this way. Hence having any positive fraction informed is not an equilibrium. Having no one informed is also not an equilibrium, because then each trader, taking the price as given, feels that there are profits to be made from becoming informed.

Efficient Markets theorists seem to be aware that costless information is a *sufficient* condition for prices to fully reflect all available information (see Fama, p. 387); they are not aware that it is a *necessary* condition. But this is a *reducto ad absurdum*, since price systems and competitive markets are important only when information is costly (see Fredrick Hayek, p. 452).

We are attempting to redefine the Efficient Markets notion, not destroy it. We have shown that when information is very inexpensive, or when informed traders get very precise information, then equilibrium exists and the market price will reveal most of the informed traders' information. However, it was argued in Section III that such markets are likely to be thin because traders have almost homogeneous beliefs.

There is a further conflict. As Grossman (1975, 1977) showed, whenever there are differences in beliefs that are not completely arbitraged, there is an incentive to create a market. (Grossman, 1977, analyzed a model of a storable commodity whose spot price did not reveal all information because of the presence of noise. Thus traders were left with differences in beliefs about the future price of the commodity. This led to the opening of a futures market. But then uninformed traders had two prices revealing information to them, implying the elimination of noise.) But, because differences in beliefs are themselves endogenous, arising out of expenditure on information and the informativeness of the price system, the creation of markets eliminates the differences of beliefs which gave rise to them, and thus causes those markets to disappear. If the creation of markets were costless, as is conventionally assumed in equilibrium analyses, equilibrium would never exist. For instance, in our model, were we to introduce an additional security, say a security which paid

$$z=\begin{cases} 1 & \text{if} \quad u>E\theta^* \\ 0 & \text{if} \quad u\leqslant E\theta^* \end{cases}$$

then the demand y for this security by the informed would depend on its price, say q, on p and on θ, while the uninformed demand depends only on p and q:

$$\lambda y_I(q,p,\theta)+(1-\lambda)y_u(q,p)=0$$

is the condition that demand equals (supply is zero for a pure security). Under weak assumptions, q and p would convey all the information concerning θ. Thus, the market would be "noiseless" and no equilibrium could exist.

Thus, we could argue as soon as the assumptions of the conventional perfect capital markets model are modified to allow even a slight amount of information imperfection and a slight cost of information, the traditional theory becomes untenable. There *cannot* be as many securities as states of nature. For if there were, competitive equilibrium would not exist.

It is only because of costly transactions and the fact that this leads to there being a limited number of markets, that competitive equilibrium can be established.

We have argued that because information is costly, prices cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain it would receive no compensation. There is a fundamental conflict between the efficiency with which markets spread information and the incentives to acquire information. However, we have said nothing regarding the social benefits of information, nor whether it is socially optimal to have "informationally efficient markets." We hope to examine the welfare properties of the equilibrium allocations herein in future work.

APPENDIX A

Here we collect some facts on conditional expectations used in the text. If X^* and Y^* are jointly normally distributed then

$$\text{(A1)}\quad E[X^*|Y^*=Y] = EX^* + \frac{Cov(X^*,Y^*)}{Var(Y^*)}\{Y-EY^*\}$$

$$\text{(A2)}\quad Var[X^*|Y^*=Y] = Var(X^*) - \frac{[Cov(X^*,Y^*)]^2}{Var(Y^*)}$$

(See Paul Hoel, p. 200.) From (A1) note that $E[X^*|Y^*]$ is a function of Y. If the expectation of both sides of (A1) is taken, we see that

$$\text{(A3)}\quad E\{E[X^*|Y^*=Y]\} = EX^*$$

Note that $Var[X^*|Y^*=Y]$ is not a function of Y, as $Var(X^*)$, $Cov(X^*,Y^*)$, and $Var(Y^*)$ are just parameters of the joint distribution of X^* and Y^*.

Two other relevant properties of conditional expectation are

$$\text{(A4)}\quad E\{E[Y^*|F(X^*)]|X^*\} = E[Y^*|F(X^*)]$$

$$\text{(A5)}\quad E\{E[Y^*|X]|F(X^*)\} = E[Y^*|F(X^*)]$$

where $F(\cdot)$ is a given function on the range of X^* (see Robert Ash, p. 260).

APPENDIX B

PROOF of Theorem 1:

(a) Suppose $\lambda=0$; then (9) becomes

$$\text{(A6)}\quad X_U(P_0(\theta,x),P_0^*) = x$$

Define

$$\text{(A7)}\quad P_0(\theta,x) \equiv \frac{E\theta^* - ax\sigma_u^2}{R}$$

where σ_u^2 is the variance of u. Note that $P_0(\theta^*,x^*)$ is uncorrelated with u^*, as x^* is uncorrelated with u^*. Hence

$$\text{(A8)}\quad E[u^*|P_0^*=P_0(\theta,x)] = Eu^* = E\theta^*$$

$$\text{and}\quad Var[u^*|P_0^*=P_0(\theta,x)] = Var[u^*]$$

Substitution of (A8) in (8) yields

$$\text{(A9)}\quad X_U(P_0^*,P_0(\theta,x)) = \frac{E\theta^* - RP_0(\theta,x)}{a\,Var\,u}$$

Substitution of (A7) in the right-hand side of (A9) yields $X_U(P_0^*(\theta,x),P_0^*)=x$ which was to be shown.

(b) Suppose $0<\lambda\leqslant 1$. Let

$$\text{(A10)}\quad P_\lambda(\theta,x) = \frac{\dfrac{\lambda w_\lambda}{a\sigma_\varepsilon^2} + \dfrac{(1-\lambda)E[u^*|w_\lambda]}{a\,Var[u^*|w_\lambda]} - Ex^*}{R\left[\dfrac{\lambda}{a\sigma_\varepsilon^2} + \dfrac{(1-\lambda)}{a\,Var[u^*|w_\lambda]}\right]}$$

Note that from equations (1), (10), (A1) and (A2):

$$\text{(A11a)}\quad E(u^*|w_\lambda) = E\theta^* + \frac{\sigma_\theta^2}{Var\,w_\lambda}\cdot(w_\lambda - E\theta^*)$$

$$\text{(A11b)}\quad Var(u^*|w_\lambda) = \sigma_\theta^2 + \sigma_\varepsilon^2 - \frac{\sigma_\theta^2}{Var\,w_\lambda}$$

$$\text{(A11c)}\quad Var\,w_\lambda = \sigma_\theta^2 + \left(\frac{a\sigma_\varepsilon^2}{\lambda}\right)^2 Var\,x^*$$

Since $P_\lambda(\theta,x)$ is a linear function of w_λ, it is immediate that $E(u^*|w_\lambda)\equiv E(u^*|P_\lambda)$, $Var(u^*|w_\lambda)=Var(u^*|P_\lambda)$, etc. To see that P_λ^* is an equilibrium, we must show that the following equation holds as an identity in (θ,x), for $P_\lambda(\cdot)$ defined by (A10):

$$\text{(A12)}\quad \lambda\cdot\frac{\theta-RP_\lambda}{a\sigma_\varepsilon^2}+(1-\lambda)\frac{E[u^*|w_\lambda]-RP_\lambda}{a\,Var[u^*|w_\lambda]}=x$$

It is immediate from (10) that (A12) holds as an identity in θ and x.

PROOF of Theorem 2:

(a) *Calculation of the expected utility of the informed.* Using the fact that W_{Ii}^λ is normally distributed conditional on $(\bar{X}_i,\theta,x)$

$$\text{(A13)}\quad E\left[V(W_{Ii}^\lambda)|\bar{X}_i,\theta,x\right]=\exp\left[-a\left\{E\left[W_{Ii}^\lambda|\bar{X}_i,\theta,x\right]-\frac{a}{2}Var\left[W_{Ii}^\lambda|\bar{X}_i,\theta,x\right]\right\}\right]$$

Using (8), (12), and the fact that (θ,x) determines a particular P,

$$\text{(A14a)}\quad E\left[W_{Ii}^\lambda|\bar{X}_i,\theta,x\right]=R(W_{0i}-c)+\frac{(E[u^*|\theta]-RP_\lambda)^2}{a\sigma_\varepsilon^2}$$

$$\text{(A14b)}\quad Var\left[W_{Ii}^\lambda|\bar{X}_i,\theta,x\right]=\frac{(E[u^*|\theta]-RP_\lambda)^2}{a^2\sigma_\varepsilon^2}$$

Substitution of (A14) into (A13) yields

$$\text{(A15)}\quad E\left[V(W_{Ii}^\lambda)|\bar{X}_i,\theta,x\right]=-\exp\left[-aR(W_{0i}-c)-\frac{1}{2\sigma_\varepsilon^2}(E[u^*|\theta]-RP_\lambda)^2\right]$$

Note that, as $P_\lambda^*(\cdot)=P_\lambda(\theta,x)$,

$$\text{(A16)}\quad E\left(E\left[V(W_{Ii}^\lambda)|\bar{X}_i,\theta,x\right]|P_\lambda,\bar{X}_i\right)=E\left[V(W_{Ii}^\lambda)|P_\lambda,\bar{X}_i\right]$$

(see (A5)). Note that since W_{0i} is nonstochastic conditional on $(P_\lambda,\bar{X}_i)$, equation (A15) implies

$$\text{(A17)}\quad E\left[V(W_{Ii}^\lambda)|P_\lambda,\bar{X}_i\right]=-\exp\left[-aR(W_{0i}^\lambda-c)\right]\cdot E\left[\left\{\exp\left[-\frac{1}{2\sigma_\varepsilon^2}(E[u|\theta]-RP_\lambda)^2\right]\right\}|P_\lambda,\bar{X}_i\right]$$

Note that by Theorem 1, conditioning on w_λ^* is equivalent to conditioning on P_λ^*. Define

$$\text{(A18)}\quad h_\lambda\equiv Var(E[u^*|\theta]|w_\lambda)=Var(\theta|w_\lambda),h_0\equiv\sigma_\varepsilon^2\equiv h$$

$$\text{(A19)}\quad Z\equiv\frac{E[u^*|\theta]-RP_\lambda}{\sqrt{h_\lambda}}$$

Using (3) and (A18), equation (A17) can be written as

$$\text{(A20)}\quad E\left[V(W_{Ii}^\lambda)|P_\lambda,\bar{X}_i\right]=e^{ac}V(RW_{0i})E\left[\exp\left[-\frac{h_\lambda}{2\sigma_\varepsilon^2}Z^2\right]|w_\lambda\right]$$

since $\bar{X}_i$ and w_λ are independent. Conditional on w_λ, P_λ is nonstochastic and $E[u^*|\theta]$ is normal. Hence conditional on $w_\lambda,(Z^*)^2$ has a noncentral *chi*-square distribution (see C. Rao, p. 181). Then for $t>0$ the moment generating function for $(Z^*)^2$ can be written

$$\text{(A21)}\quad E\left[e^{-tZ^2}|w_\lambda\right]=\frac{1}{\sqrt{1+2t}}\exp\left[\frac{-(E[Z|w_\lambda])^2t}{1+2t}\right]$$

Note that $E[u^*|\theta] = E[u^*|\theta,x]$. Hence

$$\text{(A22)}\quad E\big[E[u^*|\theta]|w_\lambda\big]=E[u^*|w_\lambda]$$
$$= E\theta^* + \frac{\sigma_\theta^2}{Var\, w_\lambda}(w_\lambda - E\theta^*)$$

since w_λ is just a function of (θ,x). Therefore

$$\text{(A23)}\quad E[Z^*|w_\lambda]=\frac{E[u^*|w_\lambda]-RP_\lambda}{\sqrt{h_\lambda}}$$

Since $u=\theta+\varepsilon$

(A24)
$$Var(u^*|w_\lambda)=\sigma_\varepsilon^2+Var(\theta^*|w_\lambda)=\sigma_\varepsilon^2+h_\lambda$$

The nondegeneracy assumptions on $(x^*, \varepsilon^*, u^*)$ imply $h_\lambda>0$. Set $t=(h_\lambda/2\sigma_\varepsilon^2)$; and evaluate (A21) using (A23) and (A24):

(A25)
$$E\left[\exp\left[-\frac{h_\lambda}{2\sigma_\varepsilon^2}Z^2\right]|w_\lambda\right]=\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}}$$
$$\cdot\exp\left(\frac{-(E(u^*|w_\lambda)-RP_\lambda)^2}{2\,Var(u^*|w_\lambda)}\right)$$

This permits the evaluation of (A20).

(b) *Calculation of expected utility of the uninformed.* Equations (8), (5), and the normality of W_{Ui}^λ conditional on w_λ can be used to show, by calculations parallel to (A13)–(A25), that

$$\text{(A26)}\quad E\big[V(W_{Ui}^\lambda)|w_\lambda,\overline{X}_i\big]$$
$$=V(RW_{0i})\exp\left(\frac{-(E(u^*|w_\lambda)-RP_\lambda)^2}{2\,Var(u^*|w_\lambda)}\right)$$

Hence

(A27)
$$E\big[V(W_{Ii}^\lambda)|w_\lambda,\overline{X}_i\big]-E\big[V(W_{Ui}^\lambda)|w_\lambda,\overline{X}_i\big]$$
$$=\left[e^{ac}\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}}-1\right]$$
$$\times E\big[V(W_{Ui}^\lambda)|w_\lambda,\overline{X}_i\big]$$

Taking expectations of both sides of (A27) yields:

$$\text{(A28)}\quad E\big[V(W_{Ii}^\lambda)\big]-E\big[V(W_{Ui}^\lambda)\big]$$
$$=\left[e^{ac}\sqrt{\frac{Var(u^*|\theta)}{Var(u^*|w_\lambda)}}-1\right]EV(W_{Ui}^\lambda)$$

Equation (13) follows immediately from (A28).

REFERENCES

Robert B. Ash, *Real Analysis and Probability,* New York 1972.

E. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance,* May 1970, *25,* 383–417.

J. R. Green, "Information, Efficiency and Equilibrium," disc. paper no. 284, Harvard Inst. Econ. Res., Mar. 1973.

______, "The Non-Existence of Informational Equilibria," *Rev. Econ. Stud.,* Oct. 1977, *44,* 451–64.

S. Grossman, "Essays on Rational Expectations," unpublished doctoral dissertation, Univ. Chicago 1975.

______, "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information," *J. Finance,* May 1976, *31,* 573–85.

______, "The Existence of Futures Markets, Noisy Rational Expectations and Informational Externalities," *Rev. Econ. Stud.,* Oct. 1977, *64,* 431–49.

______, "Further Results on the Informational Efficiency of Competitive Stock Markets," *J. Econ. Theory,* June 1978, *18,* 81–101.

______, **R. Kihlstrom, and L. Mirman,** "A Bayesian Approach to the Production of Information and Learning by Doing," *Rev. Econ. Stud.,* Oct. 1977, *64,* 533–47.

F. H. Hayek, "The Use of Knowledge in Society," *Amer. Econ. Rev.,* Sept. 1945, *35,* 519–30.

Paul G. Hoel, *Introduction to Mathematical Statistics,* New York 1962.

R. Kihlstrom and L. Mirman, "Information and Market Equilibrium," *Bell. J. Econ.,* Spring 1975, *6,* 357–76.

R. E. Lucas, Jr., "Expectations and the Neutrality of Money," *J. Econ. Theory*, Apr. 1972, *4*, 103–24.

C. Rao, *Linear Statistical Inference and Its Applications*, New York 1965.

J. E. Stiglitz, "Perfect and Imperfect Capital Markets," paper presented to the Econometric Society, New Orleans 1971.

——, "Information and Capital Markets," mimeo., Oxford Univ. 1974.

Further International Evidence on Output-Inflation Tradeoffs

By RICHARD T. FROYEN AND ROGER N. WAUD*

This paper examines new international evidence on the natural rate hypothesis in a framework originally modeled and empirically tested by Robert Lucas (1973). The Lucas model is in the spirit of the proposition also put forward by Milton Friedman, Robert Barro, and Thomas Sargent and Neil Wallace: in a nutshell, output and employment respond to aggregate demand shifts when firms and workers make errors predicting wages and prices, forming their expectations rationally, but with incomplete information. While this view has had its critics,[1] it has gained increasing attention among economists and policymakers, and even in the popular press.

Lucas's study was based on international data for the period 1952–67. Lucas himself found the data disappointing in that sixteen of the eighteen countries showed very similar behavior as measured by the variances of price and nominal aggregate demand movements. Two other countries, Argentina and Paraguay, were much more volatile. The sample thus provided "two points," Argentina and Paraguay, and the rest. The present study takes advantage of more recent data. The observations exhibit more varied price and aggregate demand behavior among countries at roughly similar developmental stages. The latter feature of the sample is desirable because the test used by Lucas relies on the assumption that several important model parameters are "relatively stable across countries" (Lucas, 1973, p. 330). Further, for a number of industrialized countries, the years since 1967 (the end point of the Lucas sample period) provide observations on a period characterized by much greater instability than for the same countries in the earlier period, at least as measured by aggregate price variance. Thus, tests of the implications of the Lucas model concerning the interrelationships among price variance, aggregate demand variance, and the terms of the output-inflation tradeoff can be made by comparing the behavior of a given country in what are apparently two regimes. The expanded opportunities for cross-country comparisons and the increased feasibility of making intertemporal comparisons in individual countries warrant a reexamination of the international evidence on output-inflation tradeoffs.

Section I of the paper reviews the Lucas model as modified by Alex Cukierman and Paul Wachtel. Section II discusses the data for this study. The results of our reexamination of the Lucas model are presented in Section III, and can be summarized as follows. The evidence supports the view that the output-inflation tradeoff deteriorates as the variance of the inflation rate increases, one implication of Lucas's model. The Lucas model also predicts an inverse relationship between the variance of the growth rate of nominal income (Lucas's measure of aggregate demand variance) and the terms of the output-inflation tradeoff. Our tests do not provide evidence to support this prediction of the model. Additionally, in a number of cases we do not find that differences in aggregate demand variance provide an adequate explanation for differences in inflation variance across countries and within a given country in different time periods, again contrary to the Lucas model's prediction. Thus our results are not consistent with the sequence hypothesized by Lucas wherein differences in inflation

---

*Associate professor and professor of economics, respectively, University of North Carolina-Chapel Hill.

[1] See, for example, William Poole's survey, particularly pp. 480–84 and Robert J. Gordon, pp. 131–32. Marcelle Arak critically reexamines the specification of Lucas's model and tests an alternative specification on *U.S.* data for the same period Lucas considered.

variance and, consequently, differing output-inflation tradeoffs are the results of differences in aggregate demand variance.

In Section IV of the paper we reexamine the assumptions of the Lucas model in the light of our empirical evidence. Possible modifications which might make the model more consistent with the data are also discussed. Section V considers the implications of our results for the natural rate hypothesis.

I. A Brief Review of the Model

The Lucas model is built on the premise that economic agents do not have enough information to distinguish relative from general price movements. As a result, suppliers of labor services and goods respond to price changes by changing quantity to a degree dependent on their past experience with the portion of price change in their market that typically represents a relative demand shift for their good or service, as opposed to a change in the general price level. In short, the greater (smaller) the variance σ^2 of the general price level relative to the variance τ^2 of the market-specific price about the mean of the general price level, the smaller (larger) will be the supply response in a specific market. When this process is aggregated across markets and agents are assumed to form expectations rationally, the model asserts that the greater the variance of the general price level relative to the variance of market-specific price changes, the less economic agents will be "fooled" into responding to nominal aggregate demand changes.

Our reexamination of the international evidence on output-inflation tradeoffs will incorporate an important modification of Lucas's model proposed in Cukierman and Wachtel. This modification introduces into the Lucas model individual equilibrium conditions for each of the markets. As a result of this modification, the variance of the market-specific price about the mean of the general price level is not constant, but will instead be a function of the variance of the economy-wide aggregate demand shock and the variance of a market-specific demand shock. Within this modified framework, the economic agent's problem is essentially unchanged. He is attempting to sort out relative price changes from aggregate price changes. The modified form of the model, however, has the advantage of consistency with the empirical evidence that the variance of relative prices is systematically related to aggregate price variance.[2]

With the above modification, the equation which gives a testable form of the Lucas hypothesis is given by

$$y_{ct} = \alpha + \pi \Delta x_t + \lambda y_{c,t-1} \tag{1}$$

where y_{ct} is the residual $(y_t - y_{nt})$ from the fitted trend line $y_{nt} = a + bt$, y_t is the *log* of real *GNP*, and Δx_t is the change in x_t, the *log* of nominal *GNP*. The parameter π measures the response of real output to a nominal aggregate demand shock, as measured by Δx_t, and λ is a speed of adjustment coefficient, $|\lambda| < 1$.

The parameter π is given by

$$\pi = \frac{\gamma}{(\sigma_x^2/\sigma_w^2) + (1+\gamma)} \tag{2}$$

where γ is a coefficient giving the supply response to an unanticipated price change (see Lucas's equation (3), 1973, p. 327), σ_w^2 is the variance of the market-specific demand shock and σ_x^2 is the variance of the nominal aggregate demand shock.[3] For fixed values of γ and σ_w^2, π will have a maximum value of $\gamma/(1+\gamma)$ at $\sigma_x^2=0$ and will decline monotonically to zero as σ_x^2 tends to infinity.

---

[2] Additionally, as Cukierman and Wachtel (fn. 3) point out, it is only with this modification that Lucas's central hypothesis concerning the relationship of the variance of nominal income change and size of the real response to such changes will follow unambiguously from the model. The model here incorporates this modification, but differs from the model in Cukierman-Wachtel in that we follow Lucas in allowing for a lagged response in the aggregate supply curve. The model in Barro also incorporates individual market-clearing conditions. Daniel Vining and Thomas Elwertowski provide evidence on the relationship between relative prices and the aggregate price level.

[3] As in the original Lucas model, the aggregate demand shock is measured by the change in nominal income (Δx), assumed to be normally distributed with mean δ and variance σ_x^2. The market-specific demand variable is a random shock whose realization differs in each market. The distribution of this random shock is assumed to be normal with mean zero and variance σ_w^2 and identical across markets. The aggregate and

The variance of the aggregate price level (σ^2) is given by

$$\sigma^2=\frac{\sigma_x^2}{(1+\theta\gamma)^2} \tag{3}$$

where $\theta=\sigma_w^2/(\sigma_x^2+\sigma_w^2)$

The closely related variance of the inflation rate will equal $2\sigma^2$. Again, assuming given values of σ_w^2 and γ, it can be shown that the variance of the inflation rate, which we denote σ_p^2, is a monotonically increasing function of σ_x^2.[4]

There are three implications that follow from the amended Lucas model if we accept the assumption that γ and σ_w^2 are relatively stable across the populations we consider:[5]

1) π and σ_x^2 should be negatively correlated (from (2));

2) σ_x^2 and σ_p^2 should be positively correlated (from (3));

3) π and σ_p^2 should be negatively correlated (from (2) and (3)).

Lucas focused exclusively on the relationship between π and σ_x^2. We will also focus on the relationships involving σ_p^2 because to do so provides some interesting insights as well as a check on the empirical consistency of the model when confronted with real world data.

II. Recent Inflationary Experience and the Data

This study is based on annual data for 1956–76 for ten industrialized countries: Belgium, Canada, France, West Germany, Italy, Japan, Netherlands, Switzerland, the United Kingdom, and the United States. In addition, quarterly data for Canada, Japan, the United Kingdom, and the United States for a comparable period were examined.[6] The most noticeable aspect of the data over the 1956–76 period is the significant increase in price variability, as measured by the estimate of σ_p^2 in many of the countries studied. This is illustrated in Table 1 where the estimate of σ_p^2 is shown for each country for the whole 1957–76 period,[7] and then for the subperiods obtained by splitting the period in half to give the estimate of σ_{p1}^2 for the 1957–66 period, and the estimate of σ_{p2}^2 for the 1967–76 period. The countries are arranged according to the size of the estimated ratio $\sigma_{p2}^2 \div \sigma_{p1}^2$ as shown in column (3) —ranging from a high of 25.105 for the United Kingdom to a low of .878 in West Germany.

While the estimated ratio $\sigma_{p2}^2 \div \sigma_{p1}^2$ is greater than one for most countries, no such pattern emerges for the estimated ratio $\sigma_{x2}^2 \div \sigma_{x1}^2$ in these countries when the subperiods are compared in this fashion, as can be seen from column (4) of Table 1. The estimated ratio $\sigma_{x2}^2 \div \sigma_{x1}^2$ is greater than one in five of the ten countries and less than one in five others, including three of the six countries (United States, Japan, and Belgium) experiencing substantial increases in inflation variability. We examine this difference in the behavior of our estimates of σ_p^2 and σ_x^2 later in the paper.

The evidence presented in Table 1 suggests that there has been a regime shift over

---

market-specific demand shocks are assumed to be independent of each other. Also note that since within this amended version (see Cukierman-Wachtel) of Lucas's model τ^2 equals $\sigma_w^2/(1+\theta\gamma)^2$ and σ^2 equals $\sigma_x^2/(1+\theta\gamma)^2$, where $\theta=\sigma_w^2/(\sigma_x^2+\sigma_w^2)$, π can be written as $\pi=\gamma/[(\sigma^2/\tau^2)+(1+\gamma)]$ which is the equivalent of Lucas's expression for π (1973, p. 330).

[4] As shown in Cukierman-Wachtel,

$$\partial\sigma_p^2/\partial\sigma_x^2=\frac{2}{(1+\theta\gamma)^3}\{1+\theta\gamma+2\theta\gamma(1-\theta)\}>0$$

where θ equals $\sigma_w^2/(\sigma_x^2+\sigma_w^2)$.

[5] Within the amended model this assumption is the equivalent of Lucas's assumption that γ and τ^2 were relatively stable across countries.

[6] The source for all the annual data is *International Financial Statistics*, May 1977, 1978. Quarterly data were kindly made available to us by the Federal Reserve Bank of St. Louis. Complete quarterly data covering the whole time period and comparable to the annual data were unavailable for the other countries in the sample. For two other countries, Germany and Belgium, data were available for one of the subperiods we will consider, as is explained below.

[7] While the data begins with 1956, the initial observation is lost due to first differencing x and p.

TABLE 1—ESTIMATE OF σ_p^2, σ_x^2 FOR 1957–76, 1957–66, 1967–76 FOR TEN COUNTRIES, ANNUAL DATA

		Whole Period 1957–76	1957–66 (1)	1967–76 (2)	$\sigma_{p2}^2 \div \sigma_{p1}^2$ (3)	$\sigma_{x2}^2 \div \sigma_{x1}^2$ (4)
United Kingdom	σ_p^2	0.00293	0.00015	0.00368	25.105	
	σ_x^2	0.00212	0.00016	0.00249		16.080
Canada	σ_p^2	0.00120	0.00011	0.00124	11.557	
	σ_x^2	0.00145	0.00068	0.00114		1.691
United States	σ_p^2	0.00054	0.00005	0.00038	7.079	
	σ_x^2	0.00059	0.00053	0.00036		0.685
Italy	σ_p^2	0.00247	0.00057	0.00297	5.195	
	σ_x^2	0.00186	0.00052	0.00242		4.670
Japan	σ_p^2	0.00142	0.00052	0.00187	3.578	
	σ_x^2	0.00161	0.00209	0.00099		0.476
Belgium	σ_p^2	0.00095	0.00030	0.00094	3.131	
	σ_x^2	0.00119	0.00073	0.00071		0.977
Switzerland	σ_p^2	0.00051	0.00024	0.00052	2.141	
	σ_x^2	0.00140	0.00067	0.00212		3.183
France	σ_p^2	0.00081	0.00061	0.00075	1.235	
	σ_x^2	0.00046	0.00034	0.00051		1.500
Netherlands	σ_p^2	0.00065	0.00042	0.00042	1.000	
	σ_x^2	0.00110	0.00137	0.00028		0.206
West Germany	σ_p^2	0.00052	0.00052	0.00046	0.878	
	σ_x^2	0.00116	0.00125	0.00106		0.853

the 1956–76 period for several countries, a finding consistent with even casual observation of events since 1967. For this reason, we examined the two regimes separately. This was done in three different ways. First, we arbitrarily split the whole period into the two equal subperiods 1957–66 and 1967–76. Second, we split the whole period in each country into the two subperiods which showed the maximum difference (in ratio form) between the size of that country's estimated price variance σ_p^2 calculated for each subperiod. Finally, we split the whole period in each country into the two subperiods showing the maximum difference between the size of the estimated nominal income variance σ_x^2 calculated for each subperiod.[8] Quarterly data for the four countries, Canada, Japan, the United Kingdom, and the United States, were examined in similar fashion.

III. The Results

Table 2 presents the estimates of π and R^2 for equation (1) for the annual data for the whole period 1957–76, and for each of the subperiods 1957–66 and 1967–76.[9] Estimates using the available quarterly data are shown in Table 3.[10] Estimates with

---

[8] The procedure for choosing the breakpoint that maximized the difference between the estimates of both σ_p^2 and σ_x^2 between subperiods was to search over all possible breakpoints subject to the constraint of having at least eight observations in each subperiod for the annual data, and twenty observations in each subperiod for the quarterly data.

[9] The complete results for the estimation of equation (1), including those for the two splits by maximum difference criteria, are given in an Appendix, Tables A:1, A:2, A:3, which is available on request from the authors.

[10] More complete quarterly estimates are given in Table A:4 of the Appendix. The estimates in Table 3 are for the 1957–75 period and the subperiods 1957–66 and 1967–75. For four countries (United States, United Kingdom, Japan, and Canada) quarterly data were available to estimate the model for a somewhat longer time period going back as far as 1953 and extending through 1976. These estimates were not significantly different from those in Table 3. A summary of the

TABLE 2—ESTIMATE OF π FOR 1957–76, 1957–66, 1967–76 FOR TEN COUNTRIES, ANNUAL DATA

Country		Whole Period 1957–76	1957–66 (1)	1967–76 (2)	$\sigma_{p2}^2+\sigma_{p1}^2$ (3)	$\sigma_{x2}^2+\sigma_{x1}^2$ (4)
United Kingdom	π	−0.119	1.207	−0.249	25.105	16.080
	t	(−1.379)	(9.176)	(−1.947)		
	R^2	0.350	0.949	0.520		
Canada	π	0.230	0.580	0.216	11.557	1.691
	t	(2.801)	(15.67)	(1.308)		
	R^2	0.656	0.990	0.276		
United States	π	0.517	0.847	1.097	7.079	0.685
	t	(2.693)	(12.77)	(2.194)		
	R^2	0.636	0.975	0.712		
Italy	π	−0.025	0.624	0.045	5.195	4.670
	t	(−0.186)	(3.233)	(0.136)		
	R^2	0.564	0.892	0.532		
Japan	π	0.493	0.519	0.399	3.578	0.476
	t	(2.634)	(6.297)	(0.768)		
	R^2	0.736	0.882	0.742		
Belgium	π	0.336	0.422	0.598	3.131	0.977
	t	(3.638)	(4.653)	(2.422)		
	R^2	0.623	0.875	0.593		
Switzerland	π	0.927	0.832	0.937	2.141	3.183
	t	(8.971)	(6.326)	(6.088)		
	R^2	0.925	0.936	0.945		
France	π	−0.122	−0.296	0.238	1.235	1.500
	t	(−0.574)	(−1.318)	(0.570)		
	R^2	0.565	0.711	0.596		
Netherlands	π	0.450	0.434	0.860	1.000	0.206
	t	(3.956)	(4.716)	(2.067)		
	R^2	0.625	0.790	0.594		
West Germany	π	0.558	0.375	0.760	0.878	0.853
	t	(4.819)	(3.668)	(3.855)		
	R^2	0.764	0.889	0.830		

quarterly data for Germany and Belgium are for the second period only, and for Belgium data were available beginning only in 1968. We focus first on cross-country comparisons and then consider intracountry comparisons over time.

A. *Cross-Country Comparisons*

According to the amended Lucas model, we should expect to find a significant negative correlation across countries between the estimated πs and σ_p^2s, and similarly a negative correlation between the estimated πs and σ_x^2s. There should be a positive correlation between the estimates of σ_p^2 and σ_x^2. For the annual data these estimated correlation coefficients for the whole period and for each of the three sets of subperiods, chosen in the manner described above, are shown in Table 4 along with their associated t-statistics. The correlation coefficients between the estimated πs and the σ_p^2s shown in the first row are statistically significant. For the time period as a whole and for the first subperiod, for each of the three data splits, the correlation coefficient between the estimates of π and σ_p^2 is significant at the 5 percent level. For the later subperiod, again for each method of dividing the period, this correlation coefficient is significant at the 1 percent level.

The other two predictions of the amended Lucas model do not receive such strong support from the data. None of the correlation coefficients between the estimates of σ_x^2

---

longer period quarterly estimates is given in Table A:5 of an Appendix, available on request from the authors.

TABLE 3—ESTIMATE OF π, σ_p^2, AND σ_x^2, 1957–75, 1957–66, 1967–75, QUARTERLY DATA

Country		Whole Period 1957–75	First Period 1957–66	Second Period 1967–75	$\sigma_{p2}^2 + \sigma_{p1}^2$	$\sigma_{x2}^2 + \sigma_{x1}^2$
Canada	σ_p^2	0.00010	0.00003	0.00011	3.667	
	σ_x^2	0.00020	0.00019	0.00015		0.789
	π	0.554	0.763	0.597		
	t	(9.283)	(14.449)	(5.321)		
	R^2	0.904	0.971	0.831		
Japan	σ_p^2	0.00017	0.00012	0.00020	1.667	
	σ_x^2	0.00037	0.00050	0.00023		0.460
	π	0.619	0.609	0.668		
	t	(9.418)	(11.609)	(4.617)		
	R^2	0.954	0.916	0.935		
United Kingdom	σ_p^2	0.00040	0.00015	0.00054	3.60	
	σ_x^2	0.00053	0.00027	0.00071		2.62
	π	0.309	0.524	0.273		
	t	(4.546)	(5.590)	(3.147)		
	R^2	0.659	0.771	0.441		
United States	σ_p^2	0.00005	0.00001	0.00004	4.000	
	σ_x^2	0.00012	0.00013	0.00009		0.692
	π	0.820	0.852	0.999		
	t	(11.770)	(30.007)	(22.139)		
	R^2	0.963	0.992	0.944		
Belgium[a]	σ_p^2			0.00013		
	σ_x^2			0.00015		
	π			0.411		
	t			(2.97)		
	R^2			.886		
West Germany	σ_p^2			0.00007		
	σ_x^2			0.00019		
	π			0.918		
	t			(10.109)		
	R^2			0.968		

[a] Beginning second quarter, 1968

TABLE 4—CORRELATIONS AMONG ESTIMATES OF π, σ_p^2, AND σ_x^2 ACROSS COUNTRIES FOR 1957–76, 1957–66, 1967–76, AND BY SUBPERIOD SPLITS
(t-statistics in parentheses)

Correlation Coefficient Between	Whole Period	Period Split in Half		Split by Maximum $\sigma_{p2}^2 - \sigma_{p1}^2$		Split by Maximum $\sigma_{x2}^2 - \sigma_{x1}^2$	
	1957–76	1957–66	1967–76	1st Regime	2d Regime	1st Regime	2d Regime
π, σ_p^2	-0.684^b	-0.685^b	-0.862^a	-0.674^b	-0.823^a	-0.690^b	-0.791^a
	(−2.653)	(−2.656)	(−4.815)	(−2.582)	(−4.095)	(−2.697)	(−3.660)
π, σ_x^2	−0.182	−0.175	-0.577^c	−0.006	−0.522	−0.030	−0.434
	(−0.524)	(−0.503)	(−1.897)	(−0.018)	(−1.731)	(−0.084)	(−1.363)
σ_x^2, σ_p^2	0.798^a	0.393	0.744^b	0.372	0.744^b	0.504	0.706^b
	(3.750)	(1.208)	(3.145)	(1.132)	(3.145)	(1.649)	(2.817)

Note: σ_p^2 = variance of change in *log* of price; σ_x^2 = variance of change in *log* of nominal income.

[a] Significant at 1 percent level.

[b] Significant at 5 percent level.

[c] Significant at 10 percent level.

and π are significant at the 5 percent level. At the 10 percent level the correlation between the estimates of σ_x^2 and π was significant, but only for the second subperiod of the arbitrary split. The correlation coefficient between σ_p^2 and σ_x^2 is significant at the 1 percent level for the period as a whole, but this correlation seems to stem from the relationship between these two variances in the second subperiod. For the earlier subperiod one cannot reject the hypothesis that the estimates of σ_p^2 and σ_x^2 are uncorrelated even at the 10 percent level. In this regard it is interesting that in the second subperiod where the estimates of σ_p^2 and σ_x^2 are more highly correlated, there is a somewhat stronger correlation between the estimates of π and σ_x^2.

Statistical cross-country comparisons of the quarterly estimates are limited by the small sample of countries, but the estimated correlation coefficients follow a pattern similar to those given in Table 4. Specifically, we find the same negative correlations between the estimates of σ_p^2 and π. For the whole period the estimated correlation coefficient is $-.91$ and for the two subperiods 1957–66 and 1967–76, the values are $-.99$ and $-.77$, respectively. All these values are significant at the 10 percent level and the value for the first subperiod is significant at the 1 percent level. Again, the relationship between the estimates of π and σ_x^2 is less strong. While the corresponding three correlation coefficients are all negative, none is significant at the 10 percent level.

B. *Intracountry Comparisons Over Time*

To facilitate comparisons of intracountry estimates over time, the countries in our sample are arranged in Table 2 according to the size of the estimated ratio of $\sigma_{p2}^2 \div \sigma_{p1}^2$.

According to the natural rate hypothesis, one would expect the output-inflation tradeoff to have worsened in countries where aggregate price variance has increased substantially. A similar pattern of decline in π estimates should be observed where σ_x^2 estimates have increased. For the countries where estimates of σ_p^2 and σ_x^2 have remained stable (or declined), π estimates should have remained stable (or increased). Finally, the estimated variances of σ_p^2 and σ_x^2 should have moved in tandem. The first six countries listed in Table 1 (United Kingdom, Canada, United States, Italy, Japan, and Belgium) are those which have experienced substantially increased inflation variance in the second subperiod.[11] The last two countries (the Netherlands and West Germany) have had stable inflation variance—in the case of West Germany a small decline. Switzerland had a relatively mild increase in inflation variance. For the other country in the sample, France, the model simply did not fit well. While we include France in any cross-country tests, we cannot discern any shifts through time in estimated π coefficients which were in all cases insignificant and generally had the wrong sign.

Three kinds of intracountry comparisons are presented. First we examine the estimates of π, σ_x^2, and σ_p^2 from the two subperiods. Second, further evidence comes from the results of Chow tests to see if one can reject the hypothesis that the set of coefficients in equation (1) is stable across the two subperiods in each country. Third, a dummy variable (d_π) was entered to test for a shift in the value of the π coefficient across the two subperiods. This variable, equal to zero in the first subperiod and one in the second, enters multiplied by Δx, and its coefficient measures the change in the π coefficient between subperiods. These estimates and their implications for the relationships among π, σ_x^2, and σ_p^2 are discussed below for the annual and quarterly data, respectively.

1. *Annual Results*

For the annual data for the subperiods obtained by the arbitrary split of the whole period, estimates are given in Table 2.[12] The

---

[11] The hypothesis $\sigma_{p1}^2 = \sigma_{p2}^2$ can be rejected at the 5 percent level for the first five countries. In Belgium the same hypothesis can be rejected at the 10 percent level. This test assumes, as did Lucas, that the distribution of the aggregate price level is normal.

[12] Complete results for estimates for subperiods split by maximum difference in price and nominal income

TABLE 5—TESTS OF COEFFICIENT STABILITY WHEN THE PERIOD IS DIVIDED BY ARBITRARY SPLIT; 1957–66, 1967–76
(t-statistics in parentheses)

Country	Chow Statistic	d_π
United Kingdom	5.09[a]	−0.088 (−0.571)
Canada	2.88[b]	−0.133 (−1.438)
United States	4.16[a]	−0.367[a] (−3.044)
Italy	0.88	−0.150 (−1.163)
Japan	0.65	−0.143 (−1.297)
Belgium	1.38	−0.089 (−0.983)
Switzerland	1.54	−0.144[b] (−1.784)
France	0.77	−0.074 (−0.830)
Netherlands	0.84	−0.062 (−0.738)
West Germany	1.97	−0.062 (−0.710)

[a]Significant at the 1 percent level.
[b]Significant at the 10 percent level.

calculated Chow statistics and the estimated coefficients for the dummy variable d_π for each of the countries are given in Table 5. As was noted in the discussion of Table 1, the positive relationship between the estimates of σ_x^2 and σ_p^2, which is predicted by the Lucas model, is not always apparent in the behavior of our estimates of σ_x^2 and σ_p^2 for the two subperiods. In three countries which experienced substantial increases in estimated σ_p^2 (United States, Belgium, and Japan), the estimate of σ_x^2 actually fell in the second subperiod. In Canada a quite large increase in estimated inflation variance (estimate of $\sigma_{p2}^2/\sigma_{p1}^2=11.557$) was accompanied by only a slight increase in the estimated nominal income variance (estimate of $\sigma_{x2}^2/\sigma_{x1}^2=1.691$). In the Netherlands there was a substantial decline in the estimate of σ_x^2 (estimate of $\sigma_{x2}^2/\sigma_{x1}^2=.206$), while the estimate of σ_{p2} was unchanged. In the other countries (United Kingdom, Italy, Switzerland, France, West Germany) estimated σ_x^2 and σ_p^2 did move together.

As for the predicted negative correlations for (σ_p^2,π) and (σ_x^2,π), the evidence from the annual data is not conclusive. The pattern of coefficient change for estimated π between the first and second subperiods is in line with the prediction on the basis of the estimated ratio $\sigma_{p2}^2 \div \sigma_{p1}^2$ for a majority of the countries in the sample. Among the countries which had substantially increased price variability, estimated π fell in four of the countries (United Kingdom, Japan, Canada, Italy). In the two price stable countries (West Germany and the Netherlands) estimated π rose in the second subperiod. For two countries (United States, Belgium) estimated π increased even though estimated σ_p^2 rose. In Switzerland estimated π also increased with a mild increase in estimated σ_p^2.

The evidence on the hypothesized negative correlation between σ_x^2 and π is not easily assessed. According to the Lucas model, a rise in σ_x^2 should decrease π *because* it will increase σ^2, the aggregate price variance (see our fn. 3), and lead agents to attribute a greater portion of the observed price change in their market to an aggregate price disturbance. Consequently, it is only for those countries where estimates of σ_x^2 and $\sigma_p^2(=2\sigma^2)$ move in the same direction that the results are consistent with the Lucas model. Among the four countries where estimated σ_x^2 and σ_p^2 did move together, in three (United Kingdom, Italy, West Germany) the π estimate moved in the expected direction. In the fourth, Switzerland, estimated σ_p^2 and σ_x^2 rose moderately and estimated π also rose, contrary to expectation.

While such an examination of the estimated coefficients from the two subperiods may be suggestive, as can be seen from the Chow statistics and the estimated coefficients on the dummy variable measuring a shift in the π coefficient, the annual data do not reveal a strong pattern of significant

---

variances appear in Tables A:2 and A:3 of an Appendix, available on request from the authors. Results with the period split according to these criteria did not differ markedly from those obtained with the arbitrary split of the data, and we do not discuss those estimates in this section.

coefficient change which would be predicted on the basis of movements in estimated σ_x^2 and σ_p^2. Among the six countries experiencing substantial increases in inflation variance, the Chow statistic is significant at the .01 level in two (United Kingdom and United States) and at the 10 percent level in one (Canada). These are the three countries with the greatest increase in inflation variance. In the other three countries experiencing substantial increases in inflation variance however, the Chow statistic is insignificant. Furthermore, the coefficient on the dummy variable testing for a shift in estimated π is significant at the 1 percent or even 5 percent level only for the United States.

2. *Quarterly Results*

For comparisons of a given country through time, the quarterly data provide more conclusive evidence on the relationship of π to σ_p^2 and σ_x^2. Estimates for the two subperiods for Canada, Japan, the United Kingdom, and the United States were given in Table 3. Table 6 shows the Chow statistics along with the estimated coefficients on the dummy variable that tests for a shift in the π coefficient between the two periods. For these four countries–all among the group experiencing substantial increases in inflation variance—the Chow test indicates that the hypothesis of coefficient stability over the two subperiods can be rejected at the 1 percent level. The estimated coefficients on the dummy variable indicate that a significant decline in the π coefficient took place in the second subperiod in all four countries. The estimated dummy variable coefficient was significant at the 1 percent level for the United States, the United Kingdom, and Canada, and at the 5 percent level for Japan.[13]

TABLE 6—CHOW STATISTICS, ESTIMATED π, AND d_π FOR QUARTERLY DATA
(t-statistics in parentheses)

Country	Chow Statistic	π	d_π[c]
United Kingdom	4.854[a]	0.603[a]	−0.338[a]
		(5.151)	(−3.043)
Canada	6.766[a]	0.754[a]	−0.268[a]
		(10.326)	(−4.086)
United States	12.932[a]	1.024[a]	−0.340[a]
		(14.471)	(−5.319)
Japan	3.634[a]	0.676[a]	−0.155[b]
		(9.969)	(−2.432)

[a]Significant at the 1 percent level.
[b]Significant at the 5 percent level.
[c]Dummy variable, π coefficient.

However the estimates from the quarterly data provide little support for the Lucas hypothesis that increases in σ_x^2 cause increases in σ_p^2 and consequent declines in π. Only in the United Kingdom is the behavior of the estimates of σ_p^2, σ_x^2, and π consistent with the predictions of the Lucas model.

IV. The Model Reconsidered

The estimates in the previous section allowed tests of three implications of the Lucas model with the following results.

1) *Estimates of π and σ_x^2 should be negatively correlated.* This proposition did not receive general support from either the cross-country tests or tests for a given country through time.

2) *Estimates of σ_x^2 and σ_p^2 should be positively correlated.* Our cross-country tests supported this proposition only in the second of the subperiods considered. Intracountry changes in estimates of σ_x^2 and σ_p^2 between subperiods did not mirror each other closely.

3) *Estimates of π and σ_p^2 should be negatively correlated.* This implication of the Lucas model was supported quite strongly by our cross-country tests. The results of the tests on individual countries through time were less conclusive. Tests with annual data did not show a pattern of statistically significant changes in estimated π coefficients for countries with substantial increases in σ_p^2. Quarterly data for four countries which did have substantial increases in σ_p^2, showed evidence of a significant decline in the

---

[13] Notice though that while the coefficient dummy variable indicates a significant drop in the π coefficient for all four countries, in two countries (United States and Japan) the point estimate for the coefficient in the second subperiod is higher than in the first.

estimated π coefficients. The quarterly data containing many more observations may provide a more powerful test for coefficient change.

Thus, our evidence does indicate that the terms of the output-inflation tradeoff deteriorate as inflation variance increases. This negative correlation between estimates of π and σ_p^2 provides support for the assumptions Lucas makes concerning the information structure governing the actions of the agents in his model. He assumes that agents cannot distinguish relative from general price movements but that they react only to perceived movements in relative prices. If agents form their expectations optimally, the higher the variance of the aggregate price level ($\sigma^2=\sigma_p^2/2$), whatever the cause, the greater the degree to which a given price change in an individual market will be taken to result from an aggregate price change. A higher aggregate price variance will result, *ceteris paribus*, in a lower real output response.

However, the results of the previous section do not provide support for the other two correlations (1 and 2 above) implied by the model. The discrepency between the model's predictions and our results calls for a reconsideration of the basic assumptions of the Lucas model, quite apart from those concerning the information structure. We will then consider possible modifications of these assumptions which might make the model more consistent with the data.

Three basic assumptions of the Lucas model are that:

A. "cyclical movements result mainly from demand shifts" (Lucas, 1977, p. 731).

B. "τ^2 and γ are relatively stable across countries" (Lucas, 1973, p. 330).

C. "nominal output is determined on the aggregate demand side of the economy" (Lucas, 1973, p. 326). We consider each in turn.

A. *Cyclical Movements Result Mainly from Demand Shifts*

The first of the assumptions listed above, that of the demand-determined nature of cyclical income movements, would seem to require some modification in light of the supply shocks which have been experienced by industrialized countries in recent years. Within the Lucas model, such supply shocks have no systematic role. They would be among the unmodeled disturbances which cause his equations to be only approximate explanations of income behavior, but, as such, cannot explain systematic differences in the behavior of economies across countries or through time.[14] Yet considering the two subperiods for our estimates, it appears that supply disturbances were more important in the second subperiod than the first, and this may explain the divergent movements of inflation variance and nominal income variance which occurred in several countries. To see this we consider a modified version of the Lucas model with an explicit aggregate supply disturbance. The effect of including a relative supply disturbance will also be discussed below. The aim of this analysis is not to provide a model which takes account of all aspects of supply shocks, but rather to suggest how adding such a supply shock to the model in a very simple way provides a possible explanation of some of our results.

Equation (4) is a modification of Lucas's supply equation for an individual market, indexed by v, which adds an explicit aggregate supply disturbance μ_t assumed to be normally distributed with zero mean,

(4)

$$y_t(v)=y_{nt}+\gamma\theta(p_t(v)-\bar{p}_t)+\lambda y_{c,t-1}(v)+\mu_t$$

where $\mu\sim N(0,\sigma_\mu^2)$. This equation, together with the demand relation,[15]

$$y_t(v)+p_t(v)=x_{t-1}+\Delta x_t+w_t(v) \tag{5}$$

where $\Delta x_t\sim N(\delta,\sigma_x^2)$ and $w_t(v)\sim N(0,\sigma_w^2)$ forms the basis for the derivation of the aggregate income equation, equivalent to equation (1) above. With the addition of the supply disturbance, assuming all three disturbances are independent of each other, we get the following expressions for π and σ_p^2

---

[14] Lucas (1977, p. 731) points out the existence of some other (but unspecified) disturbance which causes the supply curve to be less than an exact fit.

[15] Here as above we consider the Lucas model as modified by Cukierman-Wachtel.

$$\pi = \frac{\gamma}{(\sigma_\mu^2 + \sigma_x^2)/\sigma_w^2 + (1+\gamma)} \tag{2'}$$

$$\sigma_p^2 = 2\sigma^2 = \frac{2(\sigma_x^2 + \sigma_\mu^2)}{(1+\theta\gamma)^2} \tag{6}$$

with $$\theta = \frac{\sigma_w^2}{\sigma_x^2 + \sigma_\mu^2 + \sigma_w^2}$$

It follows directly that an increase in the variance of the aggregate supply disturbance will lower the real output response to an aggregate demand shock ($\partial\pi/\partial\sigma_\mu^2 < 0$). It can also be shown that such an increase in σ_μ^2 increases σ_p^2, for a given σ_x^2.[16]

Thus, the revised model would have the implication that if the more recent subperiod was characterized by more variable supply disturbances, as seems plausible, those countries experiencing such disturbances would for a given σ_x^2 have higher values of σ_p^2 and lower values of π, both the result of the rise in σ_μ^2. Such increased variance of the supply disturbance can explain the increased price variance experienced by several countries where there was no increase in estimated σ_x^2. Differences across countries in the variances of supply shocks may also provide an explanation for the first subperiod in Table 4 where there is a significant negative correlation between the π and σ_p^2 estimates, but not between the π and σ_x^2 estimates.

Other models of the natural rate hypothesis have included explicit random aggregate supply disturbances (see Sargent and Wallace, and Barro). With random disturbances to the supply curve it will be the case that the natural rate of output is itself stochastic, depending on the value taken on by the supply disturbance. Otherwise, these models are not greatly affected by the presence of a supply disturbance. As noted above, however, we do not advance the treatment here as a completely adequate way in which to model the type of supply disturbances alluded to above—poor world grain harvests, oil cartels, etc. We suggest only that the data indicate that we need to take account of such supply shocks in explaining observed changes in estimated σ_p^2 and the related changes in estimated π. The ultimate model that would result from adequately treating such supply phenomena and the resulting implications for the natural rate hypothesis remain subjects for further investigation.

Relative supply shocks can also be added to the model. If all shocks are mutually independent, an increased variance of the relative supply disturbance would reduce the variance of the inflation rate and increase the π coefficient.[17] There is some question whether supply shocks such as those from poor world grain harvests or the formation of an oil cartel should be regarded as relative or aggregate. Such supply shocks are clearly not market specific, in that they affect a broad number of markets and appear more closely approximated by aggregate shocks to supply. There may be interdependence between aggregate and market specific shocks on both the supply and demand sides of the model, broadening the question raised by Cukierman and by Vining and Elwertowski concerning causal relations between relative and aggregate shocks which we discuss in the next subsection.

B. τ^2 *and* γ *are Relatively Stable Across Countries*

In the version of the Lucas model used here, the assumption equivalent to the sec-

---

[16] The expression $\partial\sigma_p^2/\partial\sigma_\mu^2$ is identical to that given for $\partial\sigma_p^2/\partial\sigma_x^2$ in fn. 4 above, with the redefined θ.

[17] Adding a relative supply disturbance would mean putting a market indexed disturbance into equation (4). Denote such a disturbance $\eta(v) \sim N(0, \sigma_\eta^2)$, then the resulting expression for π is

$$\pi = \frac{\gamma}{(\sigma_\mu^2 + \sigma_x^2)/(\sigma_w^2 + \sigma_\eta^2) + (1+\gamma)}$$

Equation (6) remains unchanged except that now $\theta = (\sigma_w^2 + \sigma_\eta^2)/(\sigma_w^2 + \sigma_\eta^2 + \sigma_x^2 + \sigma_\mu^2)$. From these expressions it follows that $\partial\pi/\partial\sigma_\eta^2 > 0$, and $\partial\sigma_p^2/\partial\sigma_\eta^2 < 0$.

ond basic assumption is that γ, the supply response coefficient, and σ_w^2, the variance of the market-specific demand shock, are assumed to be relatively stable both across countries and for a given country over the two subperiods. Additionally, the distributions of the relative and aggregate demand disturbances (σ_w^2 and σ_x^2) are assumed to be independent. Vining and Elwertowski questioned Lucas's original assumption that σ^2 (aggregate price variance $\sigma^2 = \sigma_p^2/2$) and τ^2 were unrelated, hence, implicitly the assumption that τ^2 could be assumed constant across countries where σ^2 and σ_x^2 varied. Cukierman pointed out that a model where σ_x^2 and σ_w^2 were unrelated could explain the related changes in τ^2 and σ^2, documented by Vining and Elwertowski, if the movements in τ^2 and σ^2 both are dominated by movements in σ_x^2.

Our evidence indicates that increases in σ_x^2 do not appear to explain the increases in σ^2 that have occurred in several countries between the two subperiods we have examined (recall that $\sigma^2 = \sigma_x^2/(1+\theta\gamma)^2$ and $\theta = \sigma_w^2/(\sigma_x^2 + \sigma_w^2)$ in fn. 3), and across countries during the first subperiod. Disturbances on the supply side provide one explanation for this disparity in the behavior of the estimates of σ_p^2 and σ_x^2. Alternatively, this could be a reflection of either interdependence of the distributions of the relative and aggregate demand shocks or of systematic differences in the distribution of relative demand shocks across countries or over time. However a substantive evaluation of these possibilities is precluded by the unobservability of σ_w^2.

C. *Nominal Output is Determined on the Aggregate Demand Side of the Economy*

The last of Lucas's assumptions, that nominal income is demand determined, takes explicit form in his assumption that the aggregate demand curve is unit elastic. If aggregate demand is not unit elastic, shifts in the supply schedule will cause changes in nominal income, and therefore Δx_t in (1) cannot be assumed exogenous; it will be contemporaneously correlated with the disturbance term and this will give rise to biased estimates if (1) is estimated by ordinary least squares.[18] To check this possibility we reestimated (1) for each country and all subperiods using instrumental variables in place of Δx_t. The instruments were the money stock, exports, and the lagged value of y_{ct}. Overall, the results using instrumental variables did not suggest significant differences from the ordinary least squares results.[19]

V. Conclusion

Our reexamination of the international evidence suggests a negative correlation between aggregate price or inflation variance and the terms of the output-inflation tradeoff. Such a negative correlation is a prediction of neoclassical models, including Lucas's, which assume that agents react only to relative prices and form expectations optimally, but with only imperfect information.[20] In this sense our findings are supportive of one important element of the natural rate hypothesis.

The Lucas model also predicts a negative correlation between the terms of the output-inflation tradeoff (π) and the variance of the change in nominal income (σ_x^2) as well as a positive correlation between the variances of inflation (σ_p^2) and nominal income change (σ_x^2). Our tests provided less evidence of such correlations. In the version employed here, the Lucas hypothesis is that systematic variations across countries in both π and σ_p^2 stem predominantly from differences in σ_x^2. This sequence is not supported by the data. We find that variations in estimates of σ_x^2 do not provide a sufficient explanation of variations in estimates of π or σ_p^2 in tests both across countries or for a given country between subperiods. In Section IV we considered modifications which might make the model more consistent with the data. Modification of the model to include an aggregate supply disturbance produces a specification not inconsistent with the observed behavior of the estimates of σ_p^2, σ_x^2, and π. Interdependencies between relative and aggregate disturbances were suggested as another possible explanation of discrepancies between the model's predictions and the data. Such interdependencies may exist on either the supply or demand side. Moreover there could be interdependencies between the demand and the supply sides.

---

[18] Arak provides some evidence for the United States indicating that aggregate demand is not unit elastic and she analyzes the resulting possible bias. See also, however, Lucas's reply (1977, p. 730).

[19] The complete set of instrumental variable estimates is given in Table A:6 of an Appendix, available on request from the authors.

[20] Other examples of such neoclassical models include Barro and Lucas (1972).

REFERENCES

M. Arak, "Some International Evidence on Output-Inflation Tradeoffs: Comment," *Amer. Econ. Rev.*, Sept. 1977, *67*, 728–30.

R. J. Barro, "Rational Expectations and the Role of Monetary Policy," *J. Monet. Econ.*, Jan. 1976, *2*, 1–32.

A. Cukierman, "The Relationship Between Relative Prices and the General Price Level: A Suggested Interpretation," *Amer. Econ. Rev.*, June 1979, *69*, 444–47.

______ **and P. Wachtel,** "Differential Inflationary Expectations and the Variability of the Rate of Inflation: Theory and Evidence," *Amer. Econ. Rev.*, Sept. 1979, *69*, 595–609.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–18.

R. J. Gordon, "The Theory of Domestic Inflation," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 128–34.

R. E. Lucas, Jr., "Expectations and the Neutrality of Money," *J. Econ. Theory*, Apr. 1972, *4*, 103–24.

______, "Some International Evidence on Output-Inflation Tradeoffs," *Amer. Econ. Rev.*, June 1973, *63*, 326–34; "Errata," Dec. 1976, *66*, 985.

______, "Some International Evidence on Output-Inflation Tradeoffs: Reply," *Amer. Econ. Rev.*, Sept. 1977, *67*, 731.

W. Poole, "Rational Expectations in the Macro Model," *Brookings Papers*, Washington 1976, *2*, 463–514.

T. J. Sargent and N. Wallace, "Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *J. Polit. Econ.*, Apr. 1975, *83*, 241–54.

D. Vining and T. C. Elwertowski, "The Relationship between Relative Prices and the General Price Level," *Amer. Econ. Rev.*, Sept. 1976, *66*, 699–708.

International Monetary Fund, *Int. Finan. Statist.*, May 1977; May 1978.

Global Properties of Flexible Functional Forms

By DOUGLAS W. CAVES AND LAURITS R. CHRISTENSEN*

The economist who wishes to apply the neoclassical theory of consumer behavior to a body of data must specify a functional representation of consumer preferences.[1] The set of functional forms which are plausible candidates to represent consumer preferences has grown rapidly in the past decade. This growth is attributable to the application of duality theory to economic analysis and the concomitant development of flexible functional forms.[2] In this context flexibility is defined as the ability to represent consumer preferences without placing any prior restrictions on the full set of price and income elasticities at a base point.

Several flexible functional forms have been proposed. A partial list of these forms includes the generalized Leontief (*GL*), the translog (*TL*), the generalized Cobb-Douglas (*GCD*), the generalized square root quadratic (*GSRQ*), and the generalized Box-Cox (*GBC*).[3] Since, by definition, all these forms have the same attractive local property, it is not clear how the practitioner should choose among them. Three recent papers have discussed the issue of choosing among flexible forms. Terence Wales performed a Monte Carlo study to investigate the ability of the *GL* and *TL* forms to represent two-commodity homothetic preferences exhibiting a constant elasticity of substitution. He found that in some cases the *GL* performed better, while in other cases the *TL* performed better. Berndt, Masako Darrough, and Diewert used Canadian expenditure data to estimate three-commodity nonhomothetic *GL*, *TL*, and *GCD* forms. On the basis of better fit and conformity to neoclassical restrictions, they concluded that the *TL* form was the preferred form for their data set. Berndt and Khaled showed that the *GBC* form contains the *GL*, *GSRQ*, and *TL* forms as special or limiting cases. In their empirical example, statistical testing led to rejection of the *GSRQ* model; however, the *GL* was not rejected and results regarding the *TL* were inconclusive.

The three papers cited above employed statistical analysis as the basis for choosing among functional forms. Implicit in these statistical investigations is the recognition that the ability of a function to perform well over a range of data points, in addition to the base point, is an important criterion to be used in choosing among flexible forms. This suggests that the global properties of flexible functional forms must be known before there can be adequate understanding of the differences among these forms. To date, however, there has been no systematic theoretical analysis of the global properties of these forms. The existing evidence is based entirely upon the empirical results from specific data sets. Hence it is difficult to utilize these results to gain general insight into the properties of different flexible functional forms or to choose among forms for particular applications.

The purpose of this paper is to explore theoretically the global properties of the *GL* and *TL* forms. We confine our attention to these forms since to date they have received the widest discussion and application. The exploration provides results which indicate the range of data points over which these forms provide acceptable models of consumer behavior. We begin with two-commodity homothetic preferences. In this simple case it is feasible to provide a taxonomy of results. In more general cases we provide formulas which are applicable for any number of commodities and nonhomothetic consumer preferences. For these cases we present some examples but do not attempt to enumerate all possible results.

---

*Economics department, University of Wisconsin-Madison. We wish to thank Beth Van Zummeren for proficient research assistance. A more detailed exposition of the results in this paper is contained in discussion paper no. 7821 which is available from the Social Systems Research Institute.

[1] The issues discussed in this paper could equally well be couched in terms of the neoclassical theory of producer behavior. Rather than clutter the text with repeated references to both consumer and producer behavior, we consistently refer to the theory of consumer behavior.

[2] See W. Erwin Diewert (1974) for an excellent survey of both applications of duality theory and flexible functional forms. Lawrence Lau provides further useful discussion.

[3] The *GL*, *GCD*, and *GSRQ* forms were introduced by Diewert (1971, 1973, 1974). The *TL* was developed by Christensen, Dale Jorgenson, and Lau (1971, 1975). Ernst Berndt and Mohammed Khaled proposed the *GBC* form.

I. Empirical Implementation of the Neoclassical Theory of Consumer Behavior

The traditional starting point for the neoclassical theory of consumer behavior is the specification of a direct utility function, $U=U(X_1,\dots,X_n)$. The individual is presumed to maximize this function subject to the fixed budget constraint $Y=P_1X_1+P_2X_2+\dots+P_nX_n$. If U is monotonically increasing in the X_i and strictly quasi concave, there exists a one-to-one correspondence between the direct utility function and an indirect utility function. The indirect utility function indicates the maximum utility attainable given any set of prices and total income, $V=V((P_1/Y),(P_2/Y),\dots,(P_n/Y))$. The function V is a monotonically decreasing function of the P_i/Y and strictly quasi convex.

Since U and V are dual to each other, they provide alternative representations of any particular set of consumer preferences. Specification of the indirect utility function is generally more attractive for empirical application, since it results in a set of estimable budget share equations by simple partial differentiation. René Roy showed that the budget shares can be equated to the partial derivatives of V via the following identity:

$$(1)\qquad w_i \equiv \frac{P_iX_i}{Y} \equiv \frac{\partial V}{\partial P_i}P_i \Big/ \sum_{j=1}^{n}\frac{\partial V}{\partial P_j}P_j$$

Thus functional forms for the indirect specification of consumer preferences are of greater interest than functional forms for the direct specification. In the remainder of this paper we limit our discussion to functional forms for indirect utility functions.

Empirical implementation of Roy's Identity (1) requires that a functional form for V be chosen. Flexible functional forms are attractive because they do not place any prior restrictions on consumer behavior at a base point. A precise statement of this flexibility property follows: For any particular data point (one observation on prices and income) any set of price and income elasticities can be achieved through an appropriate set of parameter values. A nonflexible form is only capable of achieving a subset of the full range of price and income elasticities.

A flexible form can achieve arbitrary price and income elasticities at any particular data point. However, once a set of parameter values is chosen, either a priori or by statistical estimation, the pattern of price and income elasticities is determined for all possible data points. It is possible that the pattern of elasticities is plausible only for a limited range of data points, that is, that outside of a limited range of data the estimated indirect utility function is not monotonic and strictly quasi convex. In fact the range may not even include points in the sample under consideration. Particular flexible forms may differ in the range over which they provide a satisfactory representation of preferences. The extent of this capability may depend on the nature of preferences being modelled. For example, a form may perform well in cases where preferences exhibit high substitution possibilities, yet perform poorly in cases where preferences exhibit low substitution possibilities.

We are interested in determining the range over which the translog (*TL*) and generalized Leontief (*GL*) indirect utility functions provide a satisfactory representation of consumer behavior. The *TL* form can be written

$$(2)\qquad \ln V = \sum_i \alpha_i \ln(P_i/Y) + \tfrac{1}{2}\sum_i\sum_j \beta_{ij}\ln(P_i/Y)\ln(P_j/Y)$$

where $\beta_{ij}=\beta_{ji}$. The *GL* form can be written

$$(3)\quad V=\left[2\sum_i \delta_i(P_i/Y)^{1/2}+\sum_i\sum_j \gamma_{ij}(P_i/Y)^{1/2}(P_j/Y)^{1/2}\right]^{-1}$$

where $\gamma_{ij}=\gamma_{ji}$. Application of Roy's Identity to (2) results in the following budget share equations:

$$(4)\quad w_i=\left\{\alpha_i+\sum_j \beta_{ij}\,ln(P_j/Y)\right\}\Big/\left\{\sum_i \alpha_i+\sum_i\sum_j \beta_{ij}\,ln(P_j/Y)\right\}$$

Similarly, for (3) the budget share equations can be written

$$(5)\quad w_i=\left\{\delta_i(P_i/Y)^{1/2}+\sum_j \gamma_{ij}(P_i/Y)^{1/2}(P_j/Y)^{1/2}\right\}\Big/\left\{\sum_i \delta_i(P_i/Y)^{1/2}+\sum_i\sum_j \gamma_{ij}(P_i/Y)^{1/2}(P_j/Y)^{1/2}\right\}$$

The neoclassical monotonicity requirement is that the $\partial V/\partial P_i$ be strictly less than zero. From (1) it is clear that this is equivalent to requiring that the w_i be strictly greater than zero. Thus for any set of parameter values (4) and (5) can be evaluated at particular data points to determine whether monotonicity holds.

The quasi-convexity requirement on V is equivalent to the requirement that the matrix of Allen-Uzawa elasticities of substitution be negative semidefinite. Following Diewert (1974) we write the formula for the elasticities of substitution as

$$(6)\quad \sigma_{ij}=\frac{\partial^2 V}{\partial P_i\partial P_j}\left[\sum_k \frac{\partial V}{\partial P_k}P_k\right]\Big/\frac{\partial V}{\partial P_i}\frac{\partial V}{\partial P_j}-\sum_k \frac{\partial^2 V}{\partial P_j\partial P_k}P_k\Big/\frac{\partial V}{\partial P_j}-\sum_k \frac{\partial^2 V}{\partial P_i\partial P_k}P_k\Big/\frac{\partial V}{\partial P_i}+\sum_m\sum_k \frac{\partial^2 V}{\partial P_k\partial P_m}P_kP_m\Big/\sum_n \frac{\partial V}{\partial P_n}P_n$$

The matrix of elasticities of substitution is singular. Therefore, negative semidefiniteness places alternating sign restrictions on the first $n-1$ principal minors of the n-dimensional matrix.

II. Two-Commodity Homothetic Preferences

The specification of homothetic preferences is equivalent to requiring that budget shares be independent of income level, that is, that substitution possibilities be the same for all income levels. For the *TL* indirect utility function this implies that $\Sigma\alpha_i=-k$ and $\Sigma_j\beta_{ij}=0$ where k is an arbitrary positive constant. We adopt the normalization that $k=1$, which allows us to write:

$$(7)\quad V=ln\,Y+\alpha\,ln(P_1/P_2)+\frac{\beta}{2}\,ln^2(P_1/P_2)$$

$$(8)\quad w_1=-\alpha-\beta\,ln(P_1/P_2)$$

$$w_2=(1+\alpha)+\beta\,ln(P_1/P_2)$$

Monotonicity requires that w_1 and w_2 be positive. From (8) we see that w_1 and w_2 sum identically to unity. Therefore the monotonicity requirement can be stated as $0<w_1<1$. For the two-commodity case V is strictly quasi convex if and only if the elasticity of substitution (σ) between the two goods is nonnegative. For the *TL* form, σ can be written

$$(9)\qquad \sigma=(\beta/(w_1(1-w_1)))+1$$

The range of relative prices over which the regularity conditions are satisfied depends on the particular values of α and β.[4]

---

[4] We use "regularity conditions" as shorthand for monotonicity and strict quasi convexity of the indirect utility function. We use "regular region" as shorthand for "region in which the regularity conditions are satisfied."

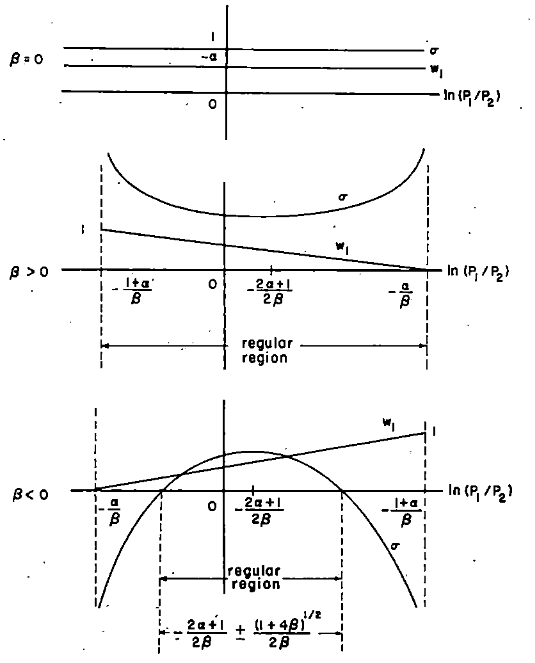


FIGURE 1. TWO-COMMODITY HOMOTHETIC TRANSLOG SHARE AND ELASTICITY OF SUBSTITUTION FUNCTIONS

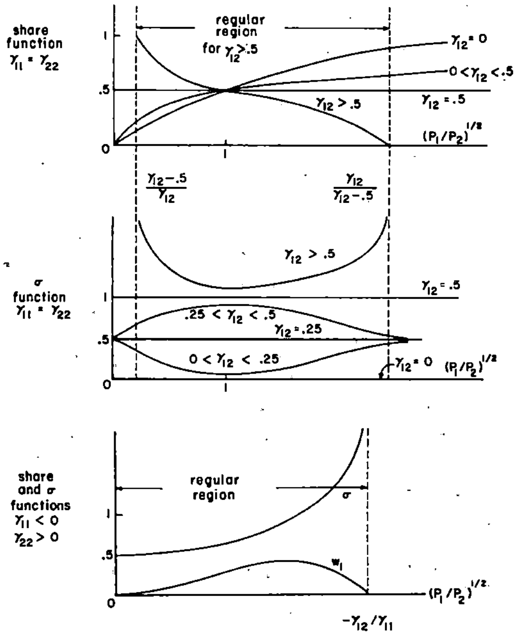


FIGURE 2. TWO-COMMODITY HOMOTHETIC GENERALIZED LEONTIEF SHARE AND ELASTICITY OF SUBSTITUTION FUNCTIONS

The three cases of interest are $\beta=0$, $\beta>0$, and $\beta<0$. Figure 1 shows the behavior of w_1 and σ for these three cases.

For $\beta=0$, the regularity conditions are satisfied for all positive prices provided that $-1<\alpha<0$. For $\beta>0$ and $\beta<0$, the size of the regular region depends on the size of β. For $\beta>0, 1+4\beta$ provides a lower bound for σ; for $\beta<0, 1+4\beta$ provides an upper bound for σ. If $\beta<-.25$, σ is negative for all price ratios—the regular region is the null set.

For the *GL* indirect utility function, homotheticity implies that $\delta_i=0$ for all i. The two-commodity homothetic *GL* indirect utility function and budget shares can be written

$$(10)\quad V=(Y/P_2)\times\{\gamma_{11}(P_1/P_2)+2\gamma_{12}(P_1/P_2)^{1/2}+\gamma_{22}\}^{-1}$$

$$(11)\quad w_1=\left(\gamma_{11}(P_1/P_2)+\gamma_{12}(P_1/P_2)^{1/2}\right)/D$$

$$w_2=\left(\gamma_{22}+\gamma_{12}(P_1/P_2)^{1/2}\right)/D$$

where $D=\gamma_{11}(P_1/P_2)+2\gamma_{12}(P_1/P_2)^{1/2}+\gamma_{22}$. The budget share equations are homogeneous of degree zero in γ_{11}, γ_{12}, and γ_{22}. Thus, without loss of generality, we can adopt the normalization $\gamma_{11}+2\gamma_{12}+\gamma_{22}=1$. For the two commodity homothetic *GL* model, the elasticity of substitution can be written

$$(12)\quad \sigma=\gamma_{12}(P_1/P_2)^{1/2}/(2w_1(1-w_1)D)$$

Since D is the ratio of income to indirect utility times P_2, it must be positive. Hence σ, w_1, and w_2 can all be nonnegative only if γ_{12} is nonnegative. There is no restriction, however, on the signs of γ_{11} and γ_{22}.

For the case $\gamma_{11}=\gamma_{22}$, the *GL* model reduces to the single parameter γ_{12}, which must be nonnegative. The first two panels of Figure 2 depict the behavior of w_1 and σ for six different values of γ_{12}. The regular region contains all positive price ratios except when $\gamma_{12}>.5$. For the case in which γ_{11} and γ_{22} are not equal but still have the same sign, the basic shapes of the share and elasticity functions are similar to those in the first two panels of Figure 2. The third panel of Figure 2 depicts the case in which γ_{11} and γ_{22} have opposite signs.

III. Indifference Curves for Two-Commodity Flexible Functional Forms

For the two-commodity case, violations of the regularity conditions can be readily interpreted in terms of the indifference curves implied by the indirect utility function. The monotonicity conditions, which are satisfied as long as both shares are positive, imply that both commodities are consumed in positive quantities. A violation of the monotonicity conditions implies that the indifference curve has crossed one of the axes. A violation of the quasi-convexity condition implies that the indifference curve is not everywhere convex to the origin.

An illuminating comparison of indifference curves from diverse functional forms can be obtained by specifying the parameters of the forms so that the indifference curves are tangent at a specified price ratio. We provide graphs of indifference curves for cases of high and low substitution possibilities. In the upper panel of Figure 3 we present *TL* and *GL* indifference curves for the case in which $\sigma=.1$ and the shares are equal to .5 at $P_1/P_2=1$. For comparison we also present an indifference curve which has a constant elasticity of substitution (*CES*) of .1 everywhere. Both the *GL* and *CES* indifference curves satisfy the regularity conditions globally; however, the *GL* curve has less curvature than the *CES* curve at price ratios other than unity. In contrast the *TL* indifference curve satisfies the regularity conditions only for a restricted region. The indifference curve exhibits a σ which declines to zero as the price ratios depart from unity. When $\sigma=0$ the indifference curve doubles back on itself. The *TL* indifference curve becomes nonconvex at the price ratios for which the σ function crosses the horizontal axis in the lower panel of Figure 1.

The lower panel of Figure 3 illustrates the relation between *TL*, *GL*, and *CES* indifference curves when preferences exhibit relatively high substitutability. In this case we have specified $\sigma=2$ at $P_1/P_2=1$. Both the *GL* and *TL* curves cross the axes, violating the monotonicity conditions at the price ratios for which the share functions become 0 or 1 in Figures 1 and 2.

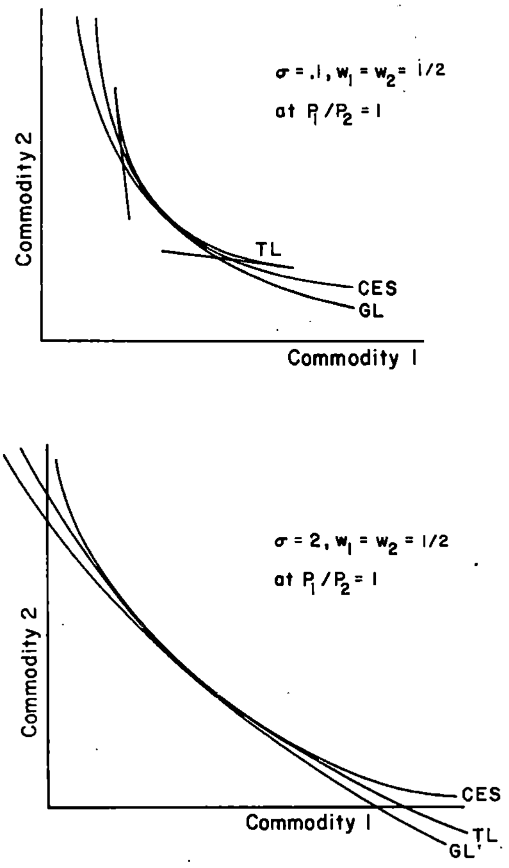


FIGURE 3. INDIFFERENCE CURVES FOR TWO-COMMODITY HOMOTHETIC PREFERENCES

IV. Multi-Commodity Homothetic Preferences

For the case of two commodities there is only a single price ratio to consider. The regular region for flexible functional forms can be described by segments of the real line. For the case of n commodities we must consider $n-1$ price ratios. The regular region can be described by regions in $n-1$ space. The monotonicity conditions require that the n shares be positive. Each w_i is a function of $n-1$ price ratios, and the region in which any w_i is positive can be shown in $n-1$ space. The intersection of n such regions is the region in which the monotonicity conditions are met. The quasi-convexity condition requires that the matrix of elasticities of substitution be negative semidefinite. This condition places restrictions on the $n-1$ principal minors of the matrix. Each of the principal minors is a function of the $n-1$ price ratios; thus the region over which the condition on any principal minor is met can be described in $n-1$ space. The intersection of these $n-1$ regions defines the region in which the quasi-convexity conditions are satisfied. Finally, the regular region is given by the intersection of the two regions defined by the monotonicity and quasi-convexity conditions.

For the n-commodity homothetic *TL* indirect utility function, the share equations can be written

$$(13) \qquad w_i = -\alpha_i - \sum_{j \neq k} \beta_{ij} \, ln(P_j/P_k)$$

The n conditions $w_i > 0$, $i = 1, \ldots, n$, which are linear in the *log* price ratios, define the region in the $n-1$ space of *log* price ratios for which the monotonicity conditions are satisfied. An example of such a region is shown in Figure 4 for the three-commodity case. The region bounded by the three straight lines shows the *log* price ratios for which the monotonicity conditions are met.

The elasticities of substitution for the n-commodity homothetic *TL* are

$$(14) \qquad \sigma_{ii} = (-\beta_{ii} - w_i + w_i^2)/w_i^2$$

$$\sigma_{ij} = -(\beta_{ij}/w_i w_j) + 1 \qquad i \neq j$$

For the three-commodity case, the quasi-convexity conditions can be written as $\sigma_{11} < 0$ and $\sigma_{11}\sigma_{22} > \sigma_{12}^2$. The region in which $\sigma_{11} < 0$ is bounded by parallel straight lines in $ln(P_1/P_3)$ and $ln(P_2/P_3)$ space. The region in which $\sigma_{11}\sigma_{22} > \sigma_{12}^2$ is bounded by a cubic function. An example of such a region is shown in Figure 4, along with the region meeting the monotonicity conditions. The triangular region in which the monotonicity conditions are satisfied is contained within the cubic region, which in turn is contained in the region bounded by the parallel lines. Thus for this example, the intersection of the regions is simply the full region in which the monotonicity conditions are satisfied.

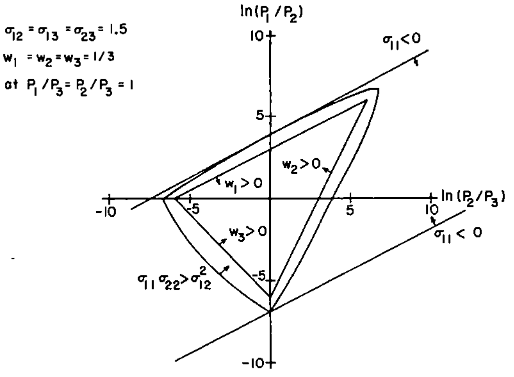


FIGURE 4. THREE-COMMODITY HOMOTHETIC TRANSLOG REGULAR REGION

For the n-commodity homothetic *GL*, the share equations are

$$(15) \qquad w_i = \left\{ \sum_j \gamma_{ij} (P_i/P_k)^{1/2} (P_j/P_k)^{1/2} \right\} / D$$

where $D = \Sigma_i \Sigma_j \gamma_{ij} (P_i/P_k)^{1/2} (P_j/P_k)^{1/2}$. Positive w_i are assured if $\Sigma_j \gamma_{ij} (P_j/P_k)^{1/2} > 0$ for all i. Thus the borders of the region in which the monotonicity conditions are met are determined by n equations, each of which is linear in the square roots of the $n-1$ price ratios. An example of such a region is shown in Figure 5 for the three-commodity case. The region bounded by the three straight lines $w_1 = 0$, $w_2 = 0$, and $w_3 = 0$ is the region in which the monotonicity conditions are met.

The elasticities of substitution for the n-commodity *GL* model are

$$(16)$$

$$\sigma_{ii} = -\tfrac{1}{2} \left\{ \sum_{j \neq i} \gamma_{ij} (P_j/P_k)^{1/2} (P_i/P_k)^{1/2} \right\} / (D w_i^2)$$

$$\sigma_{ij} = \tfrac{1}{2} \left\{ \gamma_{ij} (P_i/P_k)^{1/2} (P_j/P_k)^{1/2} \right\} / (D w_i w_j) \qquad i \neq j$$

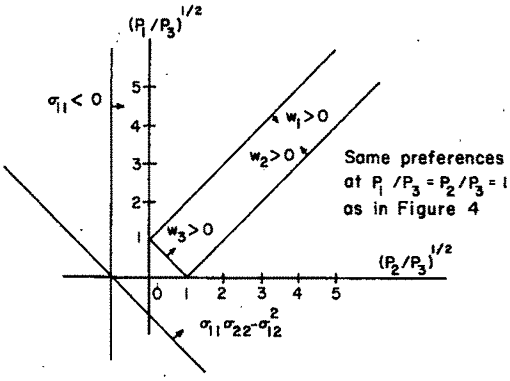


FIGURE 5. THREE-COMMODITY HOMOTHETIC GENERALIZED LEONTIEF REGULAR REGION

An example of the region in which the restrictions on the matrix of σ_{ij} are met is shown for the three-commodity case in Figure 5. The borders of the region in which quasi convexity holds are the straight lines labelled $\sigma_{11} < 0$ and $\sigma_{11}\sigma_{22} > \sigma_{12}^2$. For the example in Figure 5, the quasi-convex region includes all positive prices. As with the *TL* example, the region in which monotonicity holds also represents the intersection of the three regions.

Comparison of the regular regions in Figures 4 and 5 is difficult because of the different metrics used. In the first panel of Figure 6 we superimpose the *TL* regular region of Figure 4 on the *GL* regular region of Figure 5. For this case we find that the *GL* model can accommodate a wide range of price ratios so long as the two ratios are similar. For the *TL* neither price ratio can exceed 400, but for price ratios less than 400, the *TL* permits much greater dissimilarity in the two price ratios than the *GL*. The other five panels of Figure 6 compare *TL* and *GL* regular regions for alternative specifications of substitutability at the 1:1 price ratio.

V. Nonhomothetic Preferences

If preferences are nonhomothetic, both budget shares and substitution elasticities are dependent on the level of income as indicated by (4), (5), and (6). The regular region also becomes a function of the level of income. We illustrate this behavior for the two-commodity case.

From equation (4) we can write the two-commodity nonhomothetic *TL* share equation for w_1 as

$$\begin{aligned}(17)\quad w_1 = \big[&\alpha_1 + \beta_{11} ln(P_1/P_2) \\ &-(\beta_{11}+\beta_{12}) ln(Y/P_2)\big]/ \\ \big[&-1 + (\beta_{11}+\beta_{12}) ln(P_1/P_2) \\ &-(\beta_{11}+2\beta_{12}+\beta_{22}) ln(Y/P_2)\big]\end{aligned}$$

The corresponding elasticity of substitution can be written

$$(18)\quad \sigma = 1 - \big((w_1^{-1} w_2 \beta_{11} - 2\beta_{12} + w_2^{-1} w_1 \beta_{22})/D\big)$$

where D is the denominator of (17). Similarly, from (5) we can write the two-commodity nonhomothetic *GL* share equation for w_1 as

$$\begin{aligned}(19)\quad w_1 = \big\{&\delta_1 (P_1/P_2)^{1/2} (Y/P_2)^{1/2} \\ &+ \gamma_{11}(P_1/P_2) + \gamma_{12}(P_1/P_2)^{1/2}\big\} \\ &\times (1/D)(P_2/Y)\end{aligned}$$

where
$$\begin{aligned}D = {}&\delta_1 (P_1/P_2)^{1/2} (P_2/Y)^{1/2} \\ &+ \big(\gamma_{11}(P_1/P_2) + 2\gamma_{12}(P_1/P_2)^{1/2} + \gamma_{22}\big) \\ &\times (P_2/Y) + \delta_2 (P_2/Y)^{1/2}\end{aligned}$$

The corresponding elasticity of substitution can be written

$$\begin{aligned}(20)\quad \sigma = {}&\frac{\frac{1}{2}\gamma_{12}(P_1/P_2)^{1/2}(P_2/Y)}{D w_1 w_2} \\ &+ \frac{\frac{1}{2}\delta_1 (P_1/P_2)^{1/2}(P_2/Y)^{1/2}}{D w_1} \\ &+ \tfrac{1}{2}\delta_2 (P_2/Y)^{1/2}/D w_2 \\ &- \tfrac{1}{2}\delta_1 (P_1/P_2)^{1/2}(P_2/Y)^{1/2}/D \\ &- \tfrac{1}{2}\delta_2 (P_2/Y)^{1/2}/D\end{aligned}$$

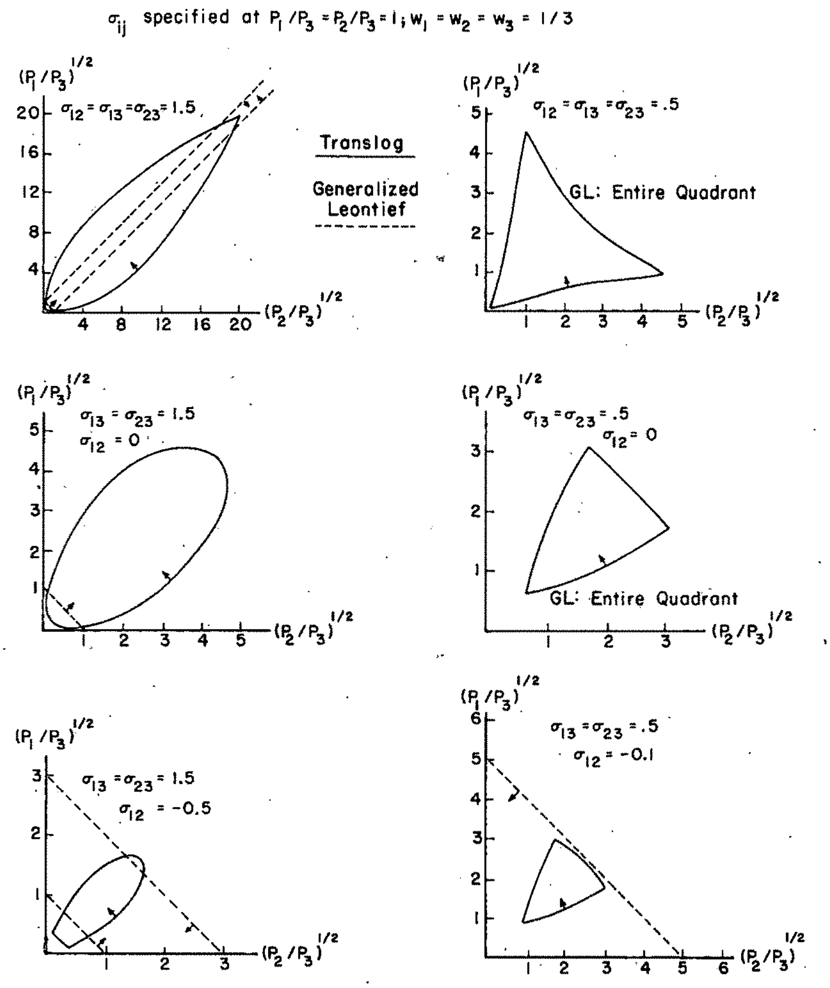


FIGURE 6. THREE-COMMODITY HOMOTHETIC TRANSLOG AND GENERALIZED LEONTIEF REGULAR REGIONS

To illustrate the regular region we must specify values for w_1, σ, the income elasticity of commodity 1 (η_{1Y}), and the change in η_{1Y} with respect to income ($\partial\eta_{1Y}/\partial Y$).[5] In Figures 7 and 8 we present comparisons of the regular regions for the *TL* and *GL* with homothetic and nonhomothetic preferences. The boundaries of the regular regions correspond to $w_1 = 0$ and $w_2 = 0$.

Figure 7 shows, for the *TL*, that the impact of income on the width of the regular region is slight when income elasticities are near unity.[6] Also, the nonhomothetic regular region is narrower except for very low levels of income. Figure 8 shows that the way in which the *GL* regular region varies with income is markedly different from the *TL*. The width of the *GL* regular region is

---

[5] The equations $w_1 + w_2 = 1$ and $w_1\eta_{1Y} + w_2\eta_{2Y} = 1$ imply values for w_2, η_{2Y}, and $\partial\eta_{2Y}/\partial Y$, given w_1, η_{1Y}, and $\partial\eta_{1Y}/\partial Y$.

[6] As *ln* (Y/P_2) changes from 0.0 to 5.0, Y/P_2 varies from 1 to 148; yet the width of the regular region shrinks by only 17 percent.

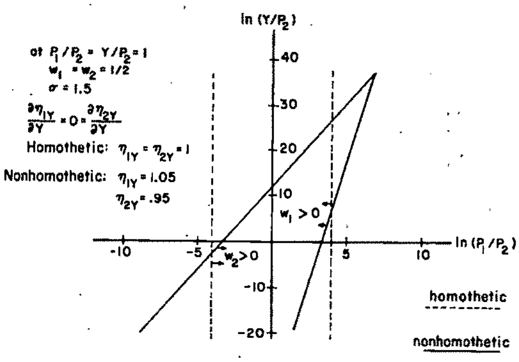
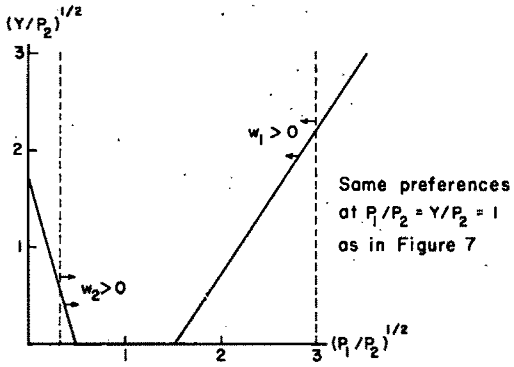


FIGURE 7. TWO-COMMODITY HOMOTHETIC AND NONHOMOTHETIC TRANSLOG REGULAR REGIONS



Same preferences at $P_1/P_2 = Y/P_2 = 1$ as in Figure 7

FIGURE 8. TWO-COMMODITY HOMOTHETIC AND NONHOMOTHETIC GENERALIZED LEONTIEF REGULAR REGIONS

σ, η_{1Y}, and η_{2Y} specified at $P_1/P_2 = Y/P_2 = 1$, $w_1 = w_2 = .5$ and $\partial \eta_{1Y}/\partial Y = 0$

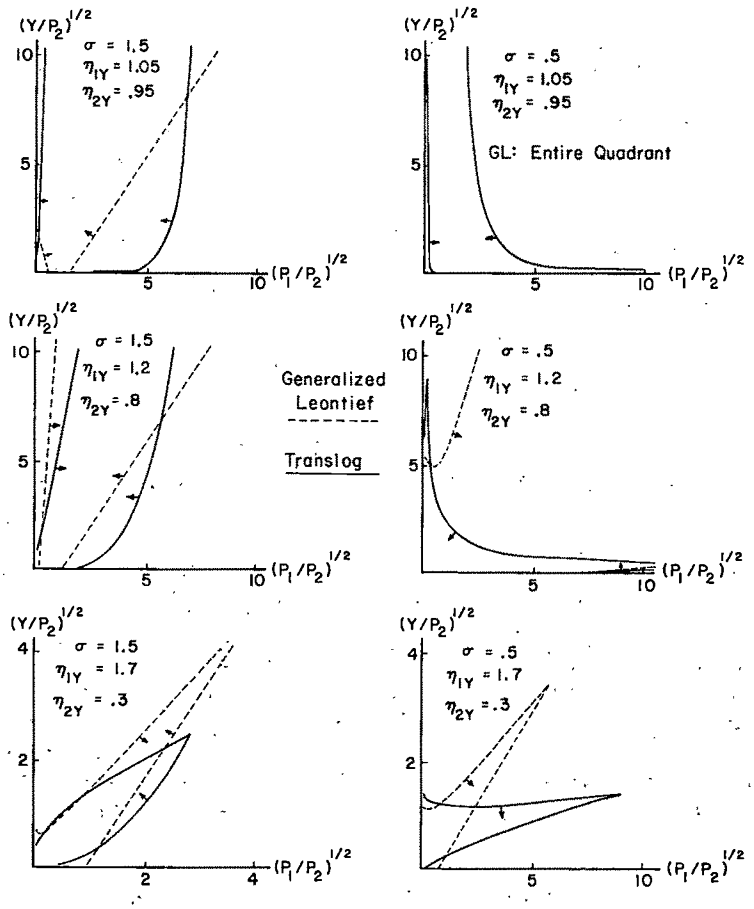


FIGURE 9. TWO-COMMODITY NONHOMOTHETIC TRANSLOG AND GENERALIZED LEONTIEF REGULAR REGIONS

positively related to income and is narrower than the homothetic region only for low levels of income.

To directly compare the *TL* and *GL* regular regions we convert the *TL* plot in Figure 7 to the square root metric and superimpose it on the *GL* region from Figure 8. The result is the first panel of Figure 9, which shows that the *TL* regular region is wider than the *GL* region, except for very large levels of income. The remaining five panels of Figure 9 present regions for other examples of nonhomothetic consumer preferences.

VI. Concluding Remarks

Given an appropriate choice of parameters, both the *GL* and the *TL* indirect utility functions can represent an arbitrary set of consumer preferences at any specified values for prices and income. However, given the parameters (either theoretically or by statistical estimation), consumer preferences are established not only for the base point but also for all possible values for prices and income. At particular points these preferences may not satisfy the restrictions of monotonicity and quasi convexity. The regions over which these regularity conditions are satisfied (regular regions) are different for the *GL* and *TL* forms. Knowledge of these regions can provide the basis for choosing between the *GL* and *TL* forms for a particular empirical application. There are two bases for choice: 1) the degree of variation in the price and income variables in the sample being considered, and 2) a priori expectations as to the magnitude of the Allen-Uzawa elasticities of substitution (*AUES*).

The *GL* and *TL* forms are not required for the analysis of two commodities with homothetic preferences. In this case they both contain two free parameters, as does the *CES* utility function (direct or indirect). The *CES* form has the advantage of satisfying the neoclassical restrictions globally for any nonnegative value of the *AUES*. The *GL* and *TL* are acceptable globally only for a subset of possible parameter values. Nonetheless, we have found it useful to display the global properties of the *GL* and *TL* forms in the two-commodity homothetic case to aid understanding of more general cases.

For more than two commodities the *CES* function is not flexible. In fact, it is impossible for any function to attain an arbitrary set of constant *AUES* (see Uzawa, 1962). Hence, although the *GL* and *TL* forms can attain an arbitrary set of *AUES* at the base point, it is foreordained that the *AUES* will vary as prices and income vary. This variation may lead to inadmissible values for particular data points. Thus, for more than two commodities the desirable flexibility property comes at the sacrifice of global satisfaction of the regularity conditions on consumer behavior. However, for certain parameter values the sacrifice is small because the *GL* and *TL* forms have very large regular regions. Furthermore, the sacrifice can be minimized by choosing the flexible form which is likely to have the larger regular region for the particular application at hand.

For two commodities and preferences which are either homothetic or nearly so, the *GL* has a larger regular region than the *TL* when the *AUES* at the base point is small. On the other hand, the *TL* is clearly preferable for high base point *AUES*. For substantial departures from homotheticity neither the *TL* nor *GL* dominate each other for high or low *AUES*.

The number of comparisons escalates rapidly as the number of commodities (n) increases. There are $n(n-1)/2$ *AUES* which are relevant. Thus comparisons in the three-commodity case require specification of three *AUES*, and comparisons in the four-commodity case require specification of six *AUES*—all in addition to specification of the income elasticities and their rates of change. We have presented a few examples for three commodities. For equal elasticities the two-commodity strengths and weaknesses of the *GL* and *TL* continue to hold. However, in cases with a mixture of high and low elasticities, the *GL* performs better than the *TL*, at least in the examples we have presented. The methods and general formulas which we have presented will allow researchers to investigate the properties of the *TL* and *GL* for any specified set of data with any desired combination of elasticities of substitution and income elasticities.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, London 1938.

E. R. Berndt, M. N. Darrough, and W. E. Diewert, "Flexible Functional Forms and Expenditure Distribution: An Application to Canadian Consumer Demand Functions," *Int. Econ. Rev.*, Oct. 1977, *18*, 651–75.

______ **and M. S. Khaled,** "Parametric Productivity Measurement and Choice Among Flexible Functional Forms," *J. Polit. Econ.*, Dec. 1979, *87*, 1220–46.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," *Econometrica*, July 1971, *39*, 255–56.

______, ______ **and** ______, "Transcendental Logarithmic Utility Functions," *Amer. Econ. Rev.*, June 1975, *65*, 367–83.

W. E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *J. Polit. Econ.*, May/June 1971, *79*, 481–507.

______, "Separability and a Generalization of the Cobb-Douglas Cost, Production, and Indirect Utility Functions," Research Branch, Department of Manpower and Immigration, Ottawa, Jan. 1973.

______, "Applications of Duality Theory," in Michael D. Intrilligator and David A. Kendrick, eds., *Frontiers of Quantitative Economics*, Vol. II, Amsterdam 1974, 176–99.

L. J. Lau, "Applications of Duality Theory: A Comment," in Michael D. Intrilligator and David A. Kendrick, eds., *Frontiers of quantitative Economics*, Vol. 2, Amsterdam 1974.

René Roy, *De L'Utilité: Contribution à la Théorie des Choix*, Paris 1943.

H. Uzawa, "Production Functions with Constant Elasticities of Substitution," *Rev. Econ. Stud.*, Oct. 1962, *29*, 291–99.

______, "Duality Principles in the Theory of Cost and Production," *Int. Econ. Rev.*, May 1964, *5*, 216–20.

T. J. Wales, "On the Flexibility of Flexible Functional Forms: An Empirical Approach," *J. Econometrics*, Mar. 1977, *5*, 183–93.

International Impact of U.S. Money Supply: The Case of Mexico

By MIKE FARRELL*

The monetary model of the balance of payments (*BOP*) implies that the long-run movements of price levels and interest rates will converge for countries which engage in trade in commodities and financial assets. Even if this implication is accepted, however, the question of interest to monetary authorities is whether deliberate policy can cause the path of domestic prices to diverge from the foreign path in the short run and, if so, how much time is required to bring about the long-run adjustment. Two papers recently published in this *Review* have contributed to discussion of this question.[1] While these models employ similar and even identical concepts, they have a fundamentally different perspective in that Mario Blejer's focus is on domestic money, income, and price variables while Lance Girton and Don Roper (hereafter, G-R) take an international or relativist view which emphasizes the role of foreign monetary policy on the *BOP*. The purpose of this paper is to present an empirical investigation of the usefulness of these two approaches in an examination of the *BOP* of Mexico, in order to evaluate the efficacy of G-R's contention that, "as far as foreign exchange market pressure is concerned, a country's monetary policy can be judged tight or easy only by reference to what is happening in the rest of the world" (p. 541).

Blejer's main objective was to relax the small country assumption employed in previous investigations of the monetary approach to the *BOP* (thus allowing the short-run divergence of prices alluded to above) by developing an ingenious framework for testing the hypothesis that the presence of nontraded goods affects the speed with which an open economy adjusts to monetary disturbances. His model explains the domestic rate of inflation and the rate of change of foreign assets held by the monetary authority, these assets being "deflated" by the total of base money outstanding. The latter variable is identical to G-R's exchange market pressure variable under a regime of fixed rates. However, under flexible rates or a mixed regime, the G-R pressure variable is a composite of the rate of change of (deflated) foreign assets of the central bank minus/plus the rate of depreciation/appreciation of the currency, the two being treated as equivalent expressions of the effect of domestic monetary policy on the demand for foreign assets. Blejer does not address the issue of exchange rate movements although his model implies that their effect on the *BOP* is manifest in their impact on the domestic price of tradeable goods. As he is forced to use the foreign inflation rate as a proxy for home prices of traded goods, he must employ a dummy variable to allow for the effect of devaluation. The significant positive coefficients of the dummy variables in his price and foreign reserves equations pick up the inflation and reserve inflows which followed Mexico's 1954 devaluation. One can only speculate as to whether, if available, a measure of domestic tradeables prices would perform as well statistically. My own preference lies with the G-R approach which makes depreciation-appreciation of the exchange a variable to be explained by the model.[2]

---

*Associate professor of economics, California State University-Long Beach. I am indebted to the Trustees of the California State University for the academic leave which allowed this research to be performed, and to the managing editor and an anonymous referee for their valuable comments and suggestions.

[1] See the papers by Mario Blejer, and by Lance Girton and Don Roper.

[2] In practice the question of how to specify the exchange rate variable is of some importance since discrete changes of parity are often large relative to the annual rate of change of foreign assets. The results

Girton and Roper employ a measure of the excess flow supply of *U.S.* base money as an explanatory variable in their study of the Canadian *BOP*. As they demonstrate, the model can be used to estimate the degree to which the monetary authorities are able to influence domestic prices and interest rates, precisely the subject to which Blejer's investigation of the role of nontraded goods is addressed. The G-R framework is more general than Blejer's, since integration of both the capital and goods markets can lead to a *BOP* offset of monetary policy, but it is not concerned formally with the question of the speed of adjustment. While the major theoretical difference among the authors lies in their choice of explanatory variables, their empirical results display one striking difference. Blejer's estimates of the *BOP* contain a large and highly significant constant term which is quite inconsistent with the long-run neutrality of money exhibited by his model. Girton and Roper on the other hand found the constant term to be small and very insignificant in their test of Canadian-*U.S.* data. As I attempt to demonstrate, it is this difference in the performance of the two models which suggests that the G-R framework is preferable.

I. Empirical Test of Girton and Roper

Since the Mexican economy is closely related to that of the United States, the G-R model is an appealing one to apply to an analysis of the Mexican *BOP*. In order to facilitate comparison of the two models it was necessary to modify slightly the original notation. The model is therefore briefly reiterated below.[3]

Assume a money demand function of the form

$$(1)\quad H_i = R_i + C_i = P_i L_i = P_i Y_i^{b_i} \exp(-\varepsilon_i \rho_i)$$

where H_i = supply of base money issued by the central bank of country i
R_i = base money created against the purchase of foreign assets
C_i = base money created by domestic credit expansion
P_i = price level
L_i = real money demand function
Y_i = real income
ρ_i = index of interest rates
b_i = income elasticity > 0
ε_i = interest rate coefficient > 0

Girton and Roper show that the monetary equilibrium condition in a two-country world can be written as

$$(2)\quad r_i - r_j + e_{ij} = -c_i + c_j + b_i y_i - b_j y_j + \theta_{ij} - \varepsilon \delta_{ij}$$

where r_i = percentage rate of change of R_i (valued in domestic currency measured against H_i of the preceding period)
e_{ij} = percentage rate of appreciation of the currency of the ith country measured in terms of the currency of the jth country
c_i = percentage rate of change of C_i measured against H_i of the preceding period
y_i = percentage rate of change of the ith country's real income
π_i = percentage rate of inflation of ith country
$\theta_{ij} = \pi_i - \pi_j + e_{ij}$ = differential rates of inflation adjusted for exchange rate changes
$\delta_{ij} = \rho_i - \rho_j$ = change in the interest differential (assuming $\varepsilon_i = \varepsilon_j$)

When one of the countries is large in terms of the size of its market for central bank liabilities, as in the case where we take country i to be Mexico and j to represent the United States, G-R rewrite equation (2)

---

reported here reflect the exchange rate measured as the annual average of daily exchange quotations, thus devaluations are weighted by the number of days of the year for which the new rate was in effect. An alternative measure which reflects devaluations entirely in the year of their occurrence actually gives a better fit. In general all the coefficients are increased in absolute value but none of them are affected significantly when this alternative is employed.

[3] As in G-R, uppercase letters represent absolute values, primes denote time rates of change, and lowercase letters show percentage rates of change of the uppercase variable.

TABLE 1

Monetary Aggregate	Coefficients of c_m	h_u	y_m	y_u	R^2	$\bar{R}^2$	$D.W$	$S.E.$
1.1 M_3	$-.95$	1.53	1.86	-1.81	.76	.73	1.91	6.03
	(.15)	(.50)	(.58)	(1.16)				
1.2 M_2	$-.82$	1.36	2.04	-2.78	.72	.68	1.92	6.56
	(.15)	(.53)	(.63)	(1.20)				
1.3 M_1	$-.82$	1.47	1.65	-2.41	.68	.64	1.76	7.08
	(.17)	(.58)	(.71)	(1.36)				

Note: The values in parentheses below the coefficients indicate the standard errors of the corresponding estimates. R^2 and $\bar{R}^2$ indicate, respectively, the coefficient of determination and the coefficient of determination adjusted for degrees of freedom. $D.W.$ is the Durbin-Watson coefficient, and $S.E.$ is the standard error of the regression. The income variables are three-year geometric lags determined as explained in the Data Appendix. All coefficients are significant at the 5 percent level or better with the exception of the coefficient of y_u in the M_3 equation which is significant at the 6.75 percent level. The regressions were constrained to pass through the origin because there is no theoretical justification for a trend term in the monetary flow equilibrium condition of equation (2).

as (substituting *m* for *i* and *u* for *j*)

(3)

$$r_m + e_m = -c_m + h_u + b_m y_m - b_u y_u + \theta_m - \varepsilon\delta_m$$

where h_u represents the percentage rate of change of *U.S.* base money. Finally, since monetary independence implies that the monetary authority is able to force the movement of the small country's price level and interest rates to diverge from those of the large country, G-R exclude the differentials of prices and interest rates from the estimating equation, arguing that the impact of the small country's monetary policy on these differentials can be inferred from the difference between the estimated coefficient of c_m and -1. The equation to be estimated is

$$(4)\quad r_m + e_m = -c_m + h_u + b_m y_m - b_u y_u + v$$

where v is a random error term.[4] Applying (4) to the data for Mexico in the period 1946–75 led to the estimates reported in Table 1.

The coefficients in Table 1 display a close resemblance to those obtained by G-R for Canada. One difference can be attributed to a variation in my procedure which was necessitated by the data and should be explained to forestall misunderstanding. The G-R approach of adjusting the monetary base to take account of policy-related changes in the money multiplier has been followed but, due to the complexity of the system of reserve requirements employed in Mexico, it was not feasible to track these changes directly. The alternative used is an indirect one in which it is assumed that the money multiplier can be represented in simplified form as the ratio of cash plus reserves to the total of cash plus bank monetary liabilities. In symbols:

$$M = N + Q = a(RM) = a(N + R)$$

where M = total money assets denominated in pesos

N = notes and coin held by the nonbank public

---

[4] The absence of a price variable is striking since (4) relates nominal money to the real demand for money. This could easily be rationalized in terms of purchasing power parity (*PPP*) since (4) assumes monetary equilibrium in both countries, i.e., $(M_i/P_i) = L_i$, and therefore $(M_m/M_u) = (P_m/P_u)(L_m/L_u)$, so that $(P_m/P_u) = (M_m/M_u)(L_u/L_m)$. Employing logarithmic differentiation, $\pi_m - \pi_u = m_m - m_u + l_u - l_m$ and thus, if $\pi_m = \pi_u$, $m_m = m_u - l_u + l_m$. G-R explicitly state that *PPP* is not assumed. Their argument is that empirically $\theta_m (= \pi_m - \pi_u + e_m)$ may be excluded from (4) because any divergence from *PPP* will be captured by the policy variables c_m and h_u while anything else which affects θ_m (real variables, commercial policy,?) will not be correlated with y_m and y_u and therefore will not bias the results.

Q = monetary liabilities of the private financial institutions
a = money multiplier
RM = reserve money
R = reserve deposits of the private financial institutions at the Bank of Mexico, and

$$a=\frac{N+Q}{N+R}=\frac{N/Q+1}{N/Q+R/Q}$$

In this framework, the policy-induced changes in the money multiplier (as opposed to those changes derived from changing preferences of the nonbank public for currency vis-à-vis bank liabilities) can be inferred from changes in the ratio of R to Q. The reserve adjustment magnitude (RAM) method of Albert Burger and Robert Rasche was employed to obtain an adjusted figure for reserve deposits which holds the ratio of R to Q constant at the level observed for 1946. Consequently the definition of M affects the measure of base money (H) through its impact on the RAM. For example, with a narrow (M_1) definition of money where Q includes only peso-denominated demand and sight deposits of the private financial institutions, the initial (1946) ratio of R to Q is .56. A somewhat broader definition (M_2) which includes time and savings deposits in Q has an initial ratio of .41 while the broadest definition (M_3) which adds the fixed-income security group of liabilities to Q has a 1946 ratio of N to Q of .37. It would of course be preferable to break these financial instruments down into money and quasi money on the basis of degrees of liquidity. Unfortunately, the fixed-income security grouping includes a range of instruments from financial bonds redeemable on demand to certificates of bank deposit which are nonnegotiable and of terms up to twenty-four months. Furthermore, in the absence of detailed reserve requirements related to each of the private financial institutions' liabilities, geographic location, etc., it is not possible to allocate reserve deposits among the various financial instruments. Thus, under a narrow definition of money, $R+RAM$ would exceed bank liabilities to the public in the years 1974 and 1975. Fortunately, as the purpose of this study is to examine the ability of the monetary authority to influence the demand for domestic base money relative to foreign reserves, use of the broadest concept of money theoretically offers the most realistic appraisal of the constraints imposed on monetary policy through the BOP—a rare case where a priori reasoning and practical constraints coincide.[5]

The estimates employing a RAM based on the M_1 and M_2 concepts of money are included in Table 1 for comparison. Not surprisingly, the offset coefficient of c_m declines with a narrower definition of money while the standard error of the regressions is increased. The major difference between Table 1 and G-R is the decline of the income coefficients and their relative weakness (in terms of significance) which is reflected in the lower R^2.

Since these results are not nearly as precise as those reported by G-R, and since we are especially interested in the model's implications for the independence of Mexico's central bank, it is useful to examine the coefficients obtained from an analysis of the full model of equation (3). This estimation is hampered by the absence of a time-series on interest rates in Mexico which covers the full period of the study. As a proxy for changes in the cost of holding money, changes in anticipated rates of inflation in Mexico and the United States were employed. These series were derived from estimates of expected inflation based on three-year lagged values of rates of change of the Mexican and *U.S.* implicit deflators

---

[5] For purposes of calculating the RAM, reserve money (RM) was considered to be the sum of nonbank cash and bank reserve deposits. The concept of base money H employed in the study proper is slightly different from RM, including, in addition to public cash and adjusted bank reserves ($N+R+RAM$), the sum of private bank holdings of currency and coin (vault cash), and nonbank public and government nonmonetary financial intermediaries' deposits at the central bank. The latter deposits have remained relatively small throughout the period under study but are included as they do represent a portion of the domestic credit extended by the Bank of Mexico.

TABLE 2

Monetary Aggregate Used	Coefficients of										
	c_m	h_u	y_m	y_u	θ_m	δ_m	R^2	$\bar{R}^2$	*D.W.*	*S.E.*	ρ
2.1a M_3	−.96	1.52	1.35	−1.19	.73	−.86	.81	.76	1.85	4.92	–
	(.19)	(.50)	(.50)	(.99)	(.23)	(.61)					
2.1b M_2	−.72	1.13	1.43	−2.11	.65	−.34	.78	.72	2.02	5.30	–
	(.16)	(.49)	(.54)	(1.00)	(.24)	(.60)					
2.1c M_1	−.79	1.30	.84	−1.30	.79	−.35	.81	.76	1.90	4.99	–
	(.15)	(.45)	(.53)	(1.00)	(.23)	(.55)					
2.2 M_3	−1.07	1.66	.91	–	.80	−1.00	.79	.75	1.73	4.98	–
	(.17)	(.49)	(.34)		(.22)	(.61)					
2.3 M_3	−.70	–	1.81	−1.25	.92	.28	.77	.73	1.88	5.33	.42
	(.17)		(.59)	(1.23)	(.25)	(.55)					(.19)
2.4 M_3	−.95	1.55	1.60	−1.33	.42	–	.80	.77	1.89	5.56	–
	(.14)	(.46)	(.55)	(1.21)	(.19)						
2.5 M_3	−1.04	1.43	1.09	–	.46	–	.79	.76	1.78	5.62	–
	(.12)	(.47)	(.36)		(.19)						
2.6 M_3	−1.04	1.99	–	1.12	.54	–	.73	.69	1.57	6.44	–
	(.16)	(.47)		(.82)	(.21)						
2.7 M_3	−.66	–	2.01	−1.76	.62	–	.72	.68	1.44	5.65	.18
	(.15)		(.54)	(1.20)	(.20)						(.20)

Note: ρ is the estimated value of the first-order autoregression coefficient. When the *D.W.* coefficient did not exceed the 1 percent upper bound for significance, the regressions were reestimated using the Cochrane-Orcutt technique to adjust for serial correlation. The *D.W.* coefficient as well as the other coefficients reflect this adjustment in those cases.

as explained in the Data Appendix. The results of these and related regressions are presented in Table 2.

Equations 2.1a, b, and c show that an increase in π_m relative to π_u has the predicted positive impact on the *BOP* and that a relative increase in the cost of holding Mexican financial assets is negatively related to r_m. Girton and Roper suggest that excluding the variables θ_m and δ_m from the regression leaves these influences to be captured by the monetary policy variables c_m and h_u. Thus in comparison with Table 1 the coefficients of c_m and h_u in equations 2.1 were expected to be larger in absolute value. (See G-R, equation (8), p. 542.) However, these results tend to show the opposite effect for the (in my opinion less reliable) M_1- and M_2-based estimates and no effect on the M_3 equation (2.1a). It may be that θ_m and δ_m reflect past changes in c_m and h_u in the estimating equation, a possibility not discussed by G-R. On the other hand, they may reflect changes in real variables which are also carried by the real income estimates y_m and y_u as suggested by the reduced size and significance of the income elasticities of the 2.1 equations. Restricting attention to the M_3 estimates, 2.2 shows that excluding the insignificant *U.S.* real income variable results in increased coefficients of the remaining variables (with the exception of y_m) and the δ_m approaches significance but with a surprisingly large coefficient. Since G-R suggested that θ_m and δ_m could replace h_u, estimate 2.3 shows the result of eliminating h_u. Mexico's monetary policy appears more effective in this version as shown by the decline in the coefficient of c_m. This may be the result of θ_m and δ_m not reflecting *U.S.* monetary policy as well as h_u does, which would at least be consistent with the possibility that these variables reflect lagged policy actions.

The instability of the coefficients of 2.1, 2.2, and 2.3 indicates the possibility of multicolinearity. In order to test for the seriousness of this problem, each of the explanatory variables was run as the regressand in a regression on all of the remaining right-hand variables. An F-test of the null hypothesis of no relationship between the regressand and the regressors was rejected at the 99 percent level in four of the six cases. Only y_m and y_u showed no significant relationship with the other five variables. The $\bar{R}^2$s of the four significant regressions ranged from .40 to .63, the latter when δ_m was the regressand.[6] Multicolinearity is clearly most serious with respect to the proxy measure of δ_m. When it is excluded from the set of explanatory variables (as in regression 2.4) and the test is repeated, none of the test regressions passes the F-test of significance at 95 percent and the $\bar{R}^2$s range from .06 to .19. The major remaining source of multicolinearity is that between y_m and y_u. When one of the income variables is the regressand, the other income coefficient has a t-value of nearly five. The effect of this interrelationship of the income variables may account for the fluctuation observed in the coefficients in Table 2 when (abandoning the proxy for δ_m) y_u, y_m, and h_u are successively excluded from the regression equation (regressions 2.5, 2.6, and 2.7, respectively). Aside from the change of sign of the (still insignificant) b_u coefficient when y_m is excluded, the most noticeable result is that the coefficient of c_m falls significantly below one only when h_u is excluded. Inclusion of the *U.S.* monetary policy variable removes the appearance of an independent capability for Mexico's central bank, thus substantiating G-R's relativist view of the *BOP*.

II. A Reexamination of the Blejer Model

Blejer's model can, with a slight alteration, be stated as follows. His analytical framework consists of two definitions,

$$(5)\qquad g \equiv \textit{Ex ante}\text{ excess supply of base money expressed as a percentage rate of change relative to } H$$

$$= c_m - \pi_m - m_d$$

where c_m and π_m have previously been defined and m_d represents the flow demand for real base money,[7] and

$$(6)\qquad P_m = \beta P_T + (1-\beta)P_{NT}$$

where P_T and P_{NT} stand for traded and nontraded goods prices while β is the share of traded goods in total expenditure. The rate of change in the relative price of traded and nontraded goods is expressed as a function of g as

$$(7)\qquad \pi_{NT} - \pi_T = \lambda(Dg)$$

which in combination with the base money flow equilibrium condition

$$(8)\qquad c_m + r_m = \pi_m + m_d$$

completes the model.

Blejer shows that the variables π_m and r_m can be expressed as functions of tradeable goods inflation, the rate of change of the excess flow supply of base money, and the prior period inflation rate π_{t-1} (the subscript m is dropped in the following).

$$(9)\qquad \pi_t = \frac{1}{1+\lambda(1-\beta)}(\pi_T)_t$$

$$+\frac{\lambda(1-\beta)}{1+\lambda(1-\beta)}(D(c-m_d)_t)+\frac{\lambda(1-\beta)}{1+\lambda(1-\beta)}\pi_{t-1}$$

---

[6] The regressions with θ_m and δ_m as regressands are

$$\theta_m = \underset{(.18)}{.29c_m} - \underset{(.48)}{.45h_u} + \underset{(.48)}{.46y_m} - \underset{(.95)}{1.10y_u} + \underset{(.46)}{1.73\delta_m}$$

$$\delta_m = -\underset{(.05)}{.20c_m} + \underset{(.15)}{.44h_u} - \underset{(.18)}{.02y_m} + \underset{(.35)}{.30y_u} + \underset{(.06)}{.24\theta_m}$$

Three of the four coefficients of c_m and h_u are significant and all have the signs predicted by the G-R model. (See G-R, equation (7), p. 542.)

[7] Blejer's concept of *ex ante* excess flow supply of money is $DC/H + a^* - (P^* + m_d^*)$ where a^* is the percentage rate of change of the money multiplier and D is the first difference operator. In this study, policy-related changes in the money multiplier are converted to changes in the domestic component of the monetary base, thus c_m is substituted for $DC/H + a^*$. The variable m_d is the rate of flow demand for real money, estimated by Blejer for several variants of the money supply. Since it hardly seems possible to isolate a domestic component of the stock of money which has its own separate demand function, I have interpreted m_d here in the G-R manner as the demand for real base money.

$$(10)\quad r_t=\frac{1}{1+\lambda(1-\beta)}(\pi_T)_t+(m_d-c)_t+\frac{\lambda(1-\beta)}{1+\lambda(1-\beta)}(D(c-m_d)_t)+\frac{\lambda(1-\beta)}{1+\lambda(1-\beta)}\pi_{t-1}$$

Equations (9) and (10) show that the effect of monetary disequilibrium on the economy can be felt on the *BOP* or on domestic prices, or a combination thereof, *depending on the structure* of the economy as reflected in the values of λ and β.

For estimation purposes π_{t-1} is eliminated from (9) and (10) by successive substitution. The reformulated expressions, writing L as the lag operator, are

$$(9')\quad \pi_t=\frac{1}{1+\lambda(1-\beta)}\sum_{i=0}^{\infty}\alpha^iL^i(\pi_T)_t+\frac{\lambda(1-\beta)}{1+\lambda(1-\beta)}\times\left(\left(\sum_{i=0}^{\infty}\alpha^iL^i-\sum_{i=0}^{\infty}\alpha^iL^{i+1}\right)(c-m_d)_t\right)$$

$$(10')\quad r_t=\frac{1}{1+\lambda(1-\beta)}\sum_{i=0}^{\infty}\alpha^iL^i(\pi_T+m_d-c)_t$$

where $\alpha=\dfrac{\lambda(1-\beta)}{1+\lambda(1-\beta)}$

Thus it is possible to estimate the current rate of inflation and the rate of "deflated" foreign reserves as functions of lagged polynomials of π_T and $c-m_d$. Equations (9′) and (10′) make it appear that the structural coefficients of the polynomials take different forms in the two estimating equations. However, further manipulation of (9′) allows the equation to be written as[8]

$$(9'')\quad \pi_t=\frac{1}{1+\lambda(1-\beta)}\sum_{i=0}^{\infty}\alpha^iL^i(\pi_T)_t+\frac{1}{1+\lambda(1-\beta)}\sum_{i=0}^{\infty}\alpha^iL^{i+1}(m_d-c)_t+\alpha(c-m_d)_t$$

When (9″) is compared to (10′) it is clear that the structural coefficients of the *lagged* variables are identical in the two equations. Apparently misled by the dissimilarity of his expressions (9′) and (10′), Blejer used the criterion of maximum R^2 to select *different* structural variables for the estimators of π_t and r_t. This *may* explain why he was able to obtain reasonably good R^2 for his explanations of these variables.

This slip could not be expected to cause the estimate of equation (10′) to be nonhomogeneous in contradiction of the underlying structural model. It can be seen from Table 3 of Blejer's paper that the best estimate of (10′) in terms of R^2 is

$$(11)\quad r_m=2.38+.51(\pi_u-X_m)+.47D_0$$

where D_0 = dummy variable for the 1954 devaluation
π_u = *U.S.* inflation used as a proxy for π_T
X_m = *ex ante* excess flow supply of base money = c_m-m_d

The perspective of the G-R model holds that it is a change in X_m relative to the excess supply of *U.S.* base money ($X_u=h_u-m_d$) which affects Mexico's foreign reserves. Thus one suspects that (11) suffers from specification error due to the omission of X_u. If true the residuals of (11) would have a nonzero expectation which would be absorbed in the constant term.[9] Furthermore the omission of X_u could lead to autocorrelation and could even bias the coefficient of π_u-X_m if X_u were correlated with that variable. As a test of these possibilities, the results of regression 1.1 were used to define new compound variables representing *ex ante* excess flow supply of money

$$(12)\quad X_m=c_m-1.86y_m$$

$$(13)\quad X_u=h_u-1.81y_u$$

---

[8] The trick is to expand several terms of the last expression in (9′) and note that $1-\alpha=1/(1+\lambda(1-\beta))$.

[9] If the true relation is $y=\beta_1X_{1i}+\beta_2X_{2i}+u_i, E(u_i)=0$ and the estimating equation is $y=\beta_1X_{1i}+V_i$ then $E(V_i)=\beta_2X_{2i}\neq 0$. Since the estimating procedure forces the sum of the residuals to zero, the omitted variable is absorbed into the constant term (see the explanation by G. S. Maddala, p. 258).

TABLE 3

	Coefficients of							
	Constant	$\pi_u^I - X_m$	$\pi_u^W - X_m$	X_u	R^2	$D.W.$	ρ	$S.E.$
3.1	−.74	.97	–	–	.67	1.18	–	6.75
	(1.34)	(.14)						
3.2	−1.18	1.05	–	–	.64	1.91	.18	6.13
	(2.20)	(.13)					(.07)	
3.3	.04	.95	–	.98	.75	1.62	–	5.97
	(1.21)	(.12)		(.35)				
3.4	−.43	–	.93	–	.65	1.44	–	6.90
	(1.37)		(.13)					
3.5	−.74	–	1.00	–	.56	1.80	.32	6.66
	(1.94)		(.15)				(.19)	
3.6	.14	–	.88	.77	.70	1.62	–	6.53
	(1.33)		(.13)	(.39)				

TABLE 4

	Coefficients of								
	Constant	π_u^I	π_u^W	X_m	X_u	R^2	$D.W.$	ρ	$S.E.$
4.1	−1.95	1.32	–	−.97	–	.68	1.21	–	6.84
	(2.39)	(.58)		(.14)					
4.2	−2.06	1.29	–	−1.05	–	.64	1.87	.43	6.25
	(3.28)	(.62)		(.13)				(.18)	
4.3	2.06	.36	–	−.94	1.20	.76	1.78	–	6.25
	(2.53)	(.61)		(.12)	(.42)				
4.4	.57	–	.64	−.97	–	.67	1.30	–	6.86
	(1.62)		(.28)	(.14)					
4.5	1.14	–	.51	−1.07	–	.61	1.93	.44	6.41
	(2.54)		(.33)	(.14)				(.18)	
4.6	2.5	–	.27	−.95	1.17	.76	1.79	–	5.91
	(1.54)		(.28)	(.12)	(.39)				

Table 3 displays the results of estimating equation (11), i.e.,

$$r_m = C_0 + C_1(\pi_m - X_m) + C_2 X_u$$

using the compound variables of (12) and (13) and two alternative measures of *U.S.* inflation, π_u^I, the implicit deflator, and π_u^W, the Wholesale Price Index (*WPI*).

The coefficient estimates of Table 3 only partially confirm the hypothesis of misspecification. When X_u is omitted, we see that $\pi_u - X_m$ is highly significant but the constant term is not. The Durbin-Watson coefficient, however, indicates the presence of autocorrelation in the residuals which is confirmed by use of the Cochrane-Orcutt technique to produce significant estimates of the first-order autocorrelation coefficient. Including X_u in the estimating equation improves the R^2 and results in no autocorrelation. It seems reasonable to infer that omitting X_u does not bias the coefficient of $\pi_u - X_m$, but does lead to serial correlation of the residuals. Contrary to speculation, the constant term is not greatly affected by omitting X_u, suggesting that the true relation is more complex than that of footnote 9.

A second possible source of specification error in equation (11) is the use of the combined variable $\pi_u - X_m$. Table 4 shows that π_u is significant when separated from X_m when it is measured by the implicit deflator but not with the *WPI*. Including X_u in the estimation removes the significance of π_u, suggesting that the *U.S.* inflation rate acts as a proxy for X_u. The large absolute

TABLE 5

		Coefficients of						
	Constant	π_u^W	θ_m	X_m	X_u^*	R^2	*D.W.*	*S.E.*
5.1	.55	.24	–	−.98	.89	.76	1.83	5.92
	(1.40)	(.28)		(.12)	(.30)			
5.2	.78	−.07	.43	−.93	.86	.80	1.91	5.53
	(1.31)	(.30)	(.21)	(.12)	(.28)			
5.3	1.09	–	–	−.96	1.01	.76	1.90	5.88
	(1.22)			(.12)	(.26)			
5.4	–	–	–	−.95	1.08	.75	1.88	5.85
				(.12)	(.25)			

size of the X_u coefficient may relate to the definition of X_u which ignores the regression coefficient of 1.5 for h_u in equation 1.1 and may therefore introduce further specification error. These points are examined in Table 5 where a new version of X_u,

$$X_u^* = 1.5h_u - 1.8y_u \tag{14}$$

is substituted for X_u and the price differential variable θ_m is added to the estimating equation.

We see from Table 5, equation 5.1, that the use of X_u^* results in a reduction of the constant term and a nonsignificant decline in its own coefficient in comparison with equation 4.6 of Table 4. Equation 5.2 shows that θ_m is significant, improving the fit of the equation, and further reducing the coefficient of π_u^W. Equations 5.3 and 5.4 show that X_m and X_u^* together can explain approximately three-quarters of the variation in r_m with coefficients quite close to their theoretical values of -1 and 1, respectively, and that it is not necessary to include a constant in the regression to obtain this result.

III. Conclusion

The relativist view of the *BOP* which emphasizes comparison of domestic money supply growth to world money supply growth is substantiated in the case of Mexico whose trade balance has long been heavily influenced by the economic conditions of the United States. However, G-R's contention that the abbreviated version of their model (equation (4)) provides a test of monetary independence does not hold up well when the full model is examined, since the price differential variable displays an additional independent role in the determination of the *BOP*. The degree to which inferences of monetary independence can be made from the estimated coefficient of the rate of change of domestic credit of the central bank appears to depend largely on the validity of the assumptions made in specifying the equilibrium conditions of the money market(s). This is a difficult econometric problem which warrants further work.

The Blejer model, which I have termed a domestic monetary model of the *BOP*, does not seem to have been appropriately tested, either in his own paper or mine. The major difficulty is the lack of data reflecting tradeable goods prices for Mexico. The use of *U.S.* prices as a proxy is unsatisfactory, since their usefulness may reflect *U.S.* monetary policy rather than the impact of goods arbitrage on Mexican prices and thus the demand for base money. For the data employed in this study, this variable is of limited value, although the result that the *U.S.* implicit deflator is better statistically than the *U.S. WPI* suggests that the direct arbitrage of goods is not an important factor in the international transmission of inflation for Mexico, and is indirect evidence that the relativist approach is preferable. An additional econometric result of this study is that the convenience of employing compound variables must be carefully weighed against the danger of specification error which can accompany the implicit restriction on the coefficients of the variables lumped together in this way.

As usual it is apparent that more work and better data are required to gain more information on how monetary pressure is exerted on the *BOP*, yet there is a clear implication that Mexico's policymakers can be advised to keep an eye on the course of the *U.S.* money supply rather than *U.S.* prices when attempting to assess the impact of their own money supply on the *BOP*.

DATA APPENDIX

Mexican international reserves were obtained by adjusting the gross international reserves figures from the Bank of Mexico as published in Bank of Mexico, *Fifty Years of Central Banking*. This series on the stock of reserves as of December 31 of each year was adjusted for *SDR* allocations and gold and foreign exchange revaluation profits, and by netting out the end-of-year dollar liabilities of the Bank of Mexico as reported in the Bank's *Annual Report*. The r_m was calculated by first differencing the resulting net reserve stock series in dollars and converting the differences to pesos with that year's exchange rate, then dividing by the lagged value of base money.

Mexican monetary base was obtained as the sum of end-of-year figures for cash and coin held by the public, deposits by the public and governmental financial institutions at the Bank of Mexico, vault cash and reserve deposits of the private banking institutions, each obtained from the Bank of Mexico's *Annual Report*, plus the *RAM*.

Reserve adjustment magnitude is equal to $(r_0 - r_t)$ *Deposits*$_t$, where

r_t = Private Bank Reserves$_t$/Private Bank Deposits$_t$, and Private Bank Deposits$_t$ were defined alternatively as

M_1: Deposits = Demand Deposits

M_2: Deposits = Demand Deposits + Time Deposits

M_3: Deposits = Demand Deposits + Time Deposits + Fixed Interest Securities

The base year $(t=0)$ is 1946.

Domestic credit of central bank of Mexico was calculated by subtracting Mexican international reserves from Mexican monetary base.

U.S. monetary base is the average of nonseasonally adjusted monthly monetary base for December and January of the following year as given in Burger and Rasche, Appendix I.

Exchange rate is the inverse of the annual average of the daily quotation for the dollar by the Bank of Mexico as given in *Fifty Years of Central Banking*.

Mexican real gross domestic product is the annual figure from Bank of Mexico, 1976.

U.S. real gross national product is the annual figure from U.S. Department of Commerce.

U.S. and Mexican price levels were measured by the implicit deflator of the real income series described above.

Expected income and expected inflation were estimated employing the method developed by Constantine Glezakos, pp. 6–8. It was assumed that expectations of future income and inflation are revised each period based on the most recent error, using the following form of partial adjustment model,

$$\text{(A1)} \qquad x_t^e = x_{t-1} + B(x_{t-1}^e - x_{t-1})$$

where the superscript e represents the expected value and x represents y or π. The reduced form of (A1), restricted to three years, is the truncated distributed lag

$$\text{(A2)} \quad x_t^e = (1-B)x_{t-1} + B(1-B)x_{t-2} + B^2(1-B)x_{t-3}$$

The adjustment coefficients B were selected by minimizing the quadratic error term

$$\text{(A3)} \qquad Q = \sum_t (x_t - x_t^e)^2$$

This was done by searching the parameter space for B from $-.9$ to $.9$ in intervals of .1, and then repeating the search at intervals of .01 in the space $\pm .1$ around the B value which minimized Q. The advantage of this procedure is that it provides a computationally simple, objective criteria for selecting the adjustment coefficients which does not rely on the R^2 of the model one is attempting to test to determine their values. The values of B employed were the following: y_m, .45; y_u, .62; π_m, .34; and π_u, .37.

REFERENCES

Robert L. Bennett, *The Financial Sector and Economic Development: The Mexican Case*, Baltimore 1965.

M. I. Blejer, "The Short-Run Dynamics of Prices and the Balance of Payments," *Amer. Econ. Rev.*, June 1977, *64*, 419–28.

Dwight S. Brothers and Leopoldo Solis, *Mexican Financial Development*, Austin 1966.

A. E. Burger and R. H. Rasche, "Revision of the Monetary Base," *Fed. Reserve Bank St. Louis Rev.*, July 1977, *59*, 13–27.

L. Girton and D. Roper, "A Monetary Model of Exchange Market Pressure Applied to the Postwar Canadian Experience," *Amer. Econ. Rev.*, Sept. 1977, *64*, 537–48.

C. Glezakos, "Inflation and Growth: A Reconsideration of the Empirical Evidence from the *LDC*'s," *J. Develop. Areas*, Jan. 1978, *12*, 171–82.

G. S. Maddala, *Econometrics*, New York 1977.

S. P. Magee, "The Empirical Evidence on the Monetary Approach to the Balance of Payments and Exchange Rates," *Amer. Econ. Rev. Proc.*, May 1976, *66*, 163–70.

Bank of Mexico, *Annual Report*, Mexico D. F., various issues.

______, *Fifty Years of Central Banking*, Mexico, D. F., 1976.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, Washington, various issues.

The Optimal Order for Submitting Manuscripts

By SHARON OSTER*

There have been a number of articles written in the last few years describing the attributes of various economics journals, including their quality, acceptance practices, and so on.[1] Presumably one could use these data both to determine which journals to buy and read, and to which journals to submit one's papers. This paper is concerned with the latter issue; in particular, I derive the optimal strategy for submitting manuscripts to economics journals.

The problem is one which we all face periodically: to which of the many available journals should a paper be sent? For highly specialized papers, the answer is sometimes obvious. In many cases, however, any one of several journals may be appropriate. This is the case dealt with in this paper.

The first problem in doing this analysis was to decide what the individual might be trying to accomplish by his or her journal choice. In general cases of this sort, one might suppose that individuals made decisions in order to maximize some kind of pecuniary return. This did not seem a fruitful way to model the individual's choice of journal. If we were really interested solely in maximizing pecuniary returns, we probably wouldn't write papers at all. In this paper, two alternative objective functions were used to represent more or less extreme, but plausible, cases: journal choice based on maximizing the stream of "prestige" points from an article versus journal choice based on maximizing the discounted stream of readers of the article. Clearly, both prestige and readership affect one's income; nevertheless these objective functions seemed to be somewhat more general than the strict rate-of-return case.

The choice process has the following characteristics. For each journal i, there is some probability, P_i, that the paper will be accepted. If it is accepted, the article earns its author a discounted stream of benefits over his or her lifetime. If the article is rejected by the first journal, it can then be submitted to a second journal, or a third. If there were no costs to a rejection, one would always adopt the strategy of first submission to the "best" journal. As it is, there is a cost to being rejected (other than the obvious psychic one): the rejection process takes time and while one is waiting the article obsolesces, tenure slots fill up with other people, and so on. In short, one trades off waiting against the quality of journal that finally accepts the article.

The formal optimization problem can be set up as follows. For each ordering of the n journals, there is a potential stream of benefits. For the ordering 1, 2, 3, 4, ..., n, for example, the benefits may be represented as:

$$(1)\quad B = P_1\int_{w_1}^{T} g_1 e^{-rt}dt + (1-P_1)P_2\int_{w_1+w_2}^{T} g_2 e^{-rt}dt, \ldots, + P_n\prod_{i=1}^{n-1}(1-P_i)\int_{\sum_{i=1}^{n} w_i}^{T} g_n e^{-rt}dt$$

where B = expected benefits from submitting to journals in the order $1,2,3\ldots n$,
$P_1\ldots P_n$ = probability of acceptance in journal $1\ldots n$
$g_1\ldots g_n$ = gain from acceptance in journal $1\ldots n$

---

*Department of economics, Yale University. I would like to thank the editorial offices of the eight journals for their help in gathering data. Laura Scher provided research assistance.

[1] Several of the most prominent articles of this genre include, Albert Danielsen and Charles Delorme; Robert Hawkins, Lawrence Ritter, and Ingo Walter (hereafter, H-R-W); Michael Lovell; Richard Quandt; Howard Tuckman and Jack Leahy; and John Weber.

$w_1 \ldots w_n$ = waiting time before decision by journal $1 \ldots n$
r = discount rate

There are $n!$ of these expected benefit streams. The problem for the individual is to choose among the $n!$ orders for the optimal order.

Several simplifications have been made in (1). I have assumed benefits end at some time T in the future. In the actual computation of (1) I assume T is retirement age, and so ignore the gain from posthumous glory. This assumption turns out to have very little empirical significance since, given the discounting, by the time retirement age comes for a current assistant professor most of the benefits have already been realized. A more significant assumption in this problem is that the Ps are independent of each other: getting rejected from the *American Economic Review* (*AER*) does not of itself diminish one's probability of acceptance in the *Journal of Political Economy* (*JPE*). The effect of time lost in the rejection process has been incorporated in the discount rate rather than through the probabilities.[2]

Eight journals were considered in this study and are listed in Table 1. The sample was limited to American journals. Other than that, the selection of journals was somewhat arbitrary, although I think there would be reasonable agreement on the list of journals chosen.

Table 1 summarizes the data used to calculate (1). The data come from several sources. The two alternative measures of gain come from H-R-W. In this survey, 160 economists were asked to rank a comprehensive list of journals on a point score of 0 to 20. Any journals with which respondents were not familiar were to be omitted. This survey generated two quality indices, which are reported in Table 1: g is the number of "prestige points" found in the H-R-W survey; f is their familiarity index. The familiarity index is my measure of the size of readership of the journal.[3] While the two indices are similar, differences do exist: a comparison of *IER* and *SEJ* is striking in this respect. Estimates of the waiting time between the submission of a manuscript and the accept/reject decision as well as the probabilities of acceptance were obtained from a telephone survey of the editorial offices of the eight journals.

Given data on r, T, P_i, g_i, and w_i ($i = 1 \ldots 8$), it is easy to compute B in (1) for any given order. Finding the optimal order (i.e., the order that maximizes B), however, is another matter. Some journals are never the optimal first choice. In particular, if we look back to Table 1, the *QJE* is dominated for all r and T by the *AER*, *Econometrica*, and the *JPE*; the *QJE* has lower prestige, longer waiting times, and lower acceptance probabilities than any of these three journals. Similarly, the *EI* is dominated by the *SEJ*. These dominance patterns were used to eliminate some of the journal permutations in calculating B. Other than that, it turned out to be easiest to simply calculate and compare all the remaining orders to find the optimal one.[4]

Given the journal characteristics listed in Table 1, the optimal strategy for submission

---

[2] Incorporating interdependence of probabilities in this problem complicates it enormously. Moreover the real connection among the probabilities is difficult to say. Suppose we order the journals 1, 2, 3...n in terms of "quality." Does a rejection from journal 1 lower our expected probability to acceptance in journal 2 more or less than a rejection from journal n? One can argue the case both ways. Suppose referee quality also varies by journal, better journals having more accurate referees. In this case one might see the rejection from journal 1 as "truer" than the rejection from journal n; P_2 would be reduced more by a rejection from journal 1 than n. Alternatively one might argue that referee accuracy is constant or randomly distributed across journals, but that standards vary—journal 1 having higher standards than journal n. In this case, P_2 would be reduced more by a rejection from n than from journal 1. Given this ambiguity in the nature of the interdependence, and in order to make the problem tractable, I assumed independence.

[3] An alternative measure of readership would be the circulation of the journal. I think the familiarity index is preferable given multiple readership of library journals.

[4] A simple Fortran program was written to perform these calculations. Computing the optimal order for each of the cases presented in Table 2 required about five minutes on the IBM 370-158 at Yale. I am indebted to Ray Fair for help with these calculations.

TABLE 1—CHARACTERISTICS OF JOURNALS

Journal	Prestige Index (1)	Familiarity Index (2)	Mean Waiting Time: Fraction of a Year (3)	Acceptance Probability (4)
American Economic Review (AER)	$g_1=18$	$f_1=100$	$w_1=.33$	$P_1=.14$
Econometrica	$g_2=16.8$	$f_2=93$	$w_2=.42$	$P_2=.22$
Economic Inquiry (EI)	$g_3=10.4$	$f_3=72$	$w_3=.19$	$P_3=.13$
International Economic Review (IER)	$g_4=13.7$	$f_4=68$	$w_4=.29$	$P_4=.25$
Southern Economic Journal (SEJ)	$g_5=12.1$	$f_5=88$	$w_5=.13$	$P_5=.19$
Journal of Political Economy (JPE)	$g_6=16.6$	$f_6=96$	$w_6=.25$	$P_6=.12$
Quarterly Journal of Economics (QJE)	$g_7=16.5$	$f_7=93$	$w_7=.5$	$P_7=.11$
Review of Economics and Statistics (REStat)	$g_8=16$	$f_8=86$	$w_8=.25$	$P_8=.17$

Source: Columns (1) and (2), Hawkins, Ritter and Walter; columns (3) and (4), editorial offices of various journals.

TABLE 2—THE OPTIMAL ORDER OF JOURNAL SUBMISSION

Case		Years to Retirement (T)	Discount Rate (r)	Benefit Index	Optimal Order[a]	Value of Benefits of Optima (B)
1.	Professor Seeking Prestige	21	.0822	Prestige: g	1,2,8,6, 4,5,7,3	108.751
2.	Assistant Professor Seeking Prestige	32	.0822	Prestige: g	1,2,8,6, 4,7,5,3	124.308
3.	Professor Seeking Readers	21	.0822	Readership: f	5,1,6,2, 8,7,3,4	622.066
4.	Assistant Professor Seeking Readers	32	.0822	Readership: f	5,1,6,2, 8,7,3,4	709.947
5.	Impatient Assistant Professor Seeking Prestige	32	.1800	Prestige: g	8,1,2,6, 4,5,7,3	54.967
6.	Patient Assistant Professor Seeking Readers	32	.0600	Readership: f	1,5,6,2, 8,7,3,4	909.881
7.	Prestige-Seeking Professor Nearing Retirement	8	.0822	Prestige: g	8,1,2,6, 4,5,7,3	58.939

[a] 1, *AER*; 2, *Econometrica*; 3, *EI*; 4, *IER*; 5, *SEJ*; 6, *JPE*; 7, *QJE*; 8, *REStat*.

will depend upon the values chosen for *r*, the discount rate, and *T*, years until retirement. I have experimented with a variety of possibilities. In Table 2, the first four cases represent the "typical" tenured (cases 1, 3) vs. untenured professor (cases 2, 4) maximizing alternatively prestige and readership. Cases 5–7 show something of the sensitivity of journal choice to changes in both expected retirement age and the discount rate.

The base-case retirement ages used were 32 years for an assistant professor and 21 years for a full professor, based on data in Tuckman and Leahy's article. The appropriate discount rate is less clear. The discount rate can be thought of as having two parts: a depreciation component representing the rate at which the knowledge embodied in the article obsolesces, and an impatience rate reflecting the anxieties of the author. Lovell has estimated a depreciation rate of 8.22 percent for economic articles; this was used as the base case for both the full and assistant professor examples.

Holding the discount rate and benefit type constant, the optimal strategy for a typical assistant professor and a typical full professor is similar. The *AER*, the highest prestige journal, is the optimal first choice for both groups; its relatively low acceptance rate and mediocre review time is not sufficient to offset its advantage in prestige points. Perhaps more surprisingly, the *SEJ* is the optimal first choice for those seeking to maximize readership. In this case, the *SEJ's* excellent review time and moderate acceptance rates push it over the top despite its middle-of-the-pack familiarity rating. As in the standard labor models, the younger economist gains more from his or her investment than the full professor, since the former has a longer life time over which the benefits may be realized.

The four base cases use a discount rate equal to the rate of obsolescence of knowledge. The typical assistant professor might be expected to be somewhat more impatient than this given the structure of the academic market. In cases 5 and 6, I have experimented with various discount rates to try to dislodge the *AER* and the *SEJ* from their front-runner positions. Both journals were strongly placed. It took a discount rate of .18 to turn the lead over to *REStat* in the prestige case; only a rate below Lovell's obsolescence rate dislodged the *SEJ*. Similarly, in case 7 I experimented with reducing *T* until the *AER* lost its lead in the readership case. As one might expect, either increasing the discount rate or radically shortening expected life time favors those journals, like *REStat* and *SEJ*, characterized by relatively high acceptance rates and quick reviews. It appears that, at least in the context of this problem, growing old is rather like becoming an assistant professor again.

Is there any evidence that economists behave as if they were maximizing equation (1)? Full experimental confirmation would require a detailed psychological and demographic profile of all authors who submitted manuscripts to the various journals, along with the order of submission. I offer instead some casual evidence, more easily obtained. The results in Table 2 suggest that very young, impatient assistant professors should prefer submitting manuscripts to either *REStat* or *SEJ* over the *AER*. If the three journals have similar patterns of acceptances across the different age groups, then we should observe a higher frequency of young economists in *REStat* and *SEJ*. I have calculated for 1977 the percent of articles in the three journals written by economists born after 1947 (30 was rather arbitrarily chosen as young).[5] In fact, as predicted, the *AER* had the lowest share of articles from this group: 17.1 percent as opposed to 25.9 percent for the *SEJ* and 38.1 percent for *REStat*. These patterns may, of course, be reinforced once this article makes apparent to some their inappropriate submission practices.

It is easy to generalize these results using other discount rates, measures of gain, and so on. In fact, the computational problem is easy enough so that the optimal order should always be calculated. Finally, as I am sure it will be asked, I have calculated the optimal order for me, a slightly impatient, prestige seeking assistant professor ($r=.15; T=32$). My optimal order is *AER*, *REStat*, *Econometrica*, *JPE*, *IER*, *SEJ*, *QJE*, *EI*. So given the journal you are now reading, you know how many times this paper was rejected.

---

[5] Only *AER* members were counted since the *1974 Directory of Members* was used to determine ages.

REFERENCES

A. Danielsen and C. Delorme, "Some Empirical Evidence on the Variables Associated with the Ranking of Economics Journals," *Southern Econ. J.*, Oct. 1976, *43*, 1149–60.

R. Hawkins, L. Ritter, and I. Walter, "What Economists Think of Their Journals," *J. Polit. Econ.*, July/Aug. 1973, *81*, 1017–32.

M. Lovell, "The Production of Economic Literature: An Interpretation," *J. Econ. Lit.*, Mar. 1973, *11*, 27–55.

R. Quandt, "Some Quantitative Aspects of the Economics Journal Literature," *J. Polit. Econ.*, Aug. 1976, *84*, 741–56.

H. Tuckman and J. Leahy, "What is an Article Worth?," *J. Polit. Econ.*, Oct. 1975, *83*, 951–68.

J. Weber, "Economics Journals: Policies, Funds and Problems," *Southern Econ. J.*, Apr. 1972, *38*, 559–65.

American Economic Association, *1974 Directory of Members*, Oct. 1974, Suppl., *64*, No. 5.

Consumer's Surplus in Commodity Space

By ALAN RANDALL AND JOHN R. STOLL*

In an article which is already widely quoted, Robert Willig demonstrated that observed Marshallian consumer's surplus can be rigorously utilized to estimate the (theoretically correct) Hicksian compensating and equivalent variations. Willig's analysis was confined to price space, and thus finds its application in the use of consumer's surplus to evaluate the welfare impacts of price changes.

There is another important area of empirical research in which consumer's surplus finds application: the evaluation of the benefits and costs of proposed projects and programs.[1] In benefit-cost analysis, the economist is often concerned not so much with the welfare impacts of price changes as with the welfare impacts of changes in the bundle of goods, services, and amenities possessed, used, or consumed by individuals. Proposed projects and programs may remove some goods from individual opportunity sets, while introducing new goods; and may decrease the quantities of some goods while increasing the quantities of others. Typically, these changes in individually held bundles of specific goods are not accompanied by directly related and commensurate changes in income (or the value of "all other goods").

The goods affected by proposed projects and programs may include divisible, exclusive goods whose competitive prices can be observed in well-functioning markets. They may also include recreational and environmental amenities, the existence of unique environments and endangered species, and increases or decreases in expected human mortality or morbidity, etc.: goods which may be in varying degrees indivisible, nonexclusive, and unpriced.

The concern which motivated Willig's article, in particular, that observable consumer's surplus data are usually not in the theoretically correct form, apply to an even greater extent to the estimation of the welfare impacts of changes in the bundle of goods. In the case of priced goods, only Marshallian consumer's surplus is directly observable; for unpriced goods, value data may, depending on the estimation technique used, be in the form of Marshallian consumer's surplus, the Hicksian compensating measure, the Hicksian equivalent measure, or the expenditure function. This paper identifies the conditions under which Willig's conclusion that consumer's surplus data may be rigorously used "without apology" is adaptable to situations in which it is bundles of goods, rather than prices, which are changed.

Our findings, in broad outline, are as follows. (a) For goods which are perfectly divisible and exchanged at zero transactions costs in infinitely large markets, the Hicksian compensating and equivalent variation measures of consumer's surplus are equal in absolute value and are equal to the price multiplied by the quantity change. (b) For goods which are indivisible or lumpy, the Hicksian compensating measure of welfare loss (gain) is larger (smaller) in absolute value than the Marshallian consumer's surplus measure, which in turn is larger (smaller) than the Hicksian equivalent measure.[2] Given information on the price flexibility of income for the good in question, the individual's income and the proportion of that income which is spent on the good, bounds on the difference between the

---

*University of Kentucky and Texas A&M University, respectively. This paper (no. 79-1-3) is published with the approval of the Director of the Kentucky Agricultural Experiment Station. We appreciate helpful comments from an anonymous reviewer.

[1] See E. J. Mishan for example.

[2] This finding is in contrast to that of Karl-Göran Mäler (pp. 131-40) which has been cited without question by such authors as Charles Cicchetti, Anthony Fisher, and V. Kerry Smith.

compensating and equivalent measures can be rigorously calculated. (c) For divisible goods traded in costly markets (i.e., those where transactions costs are positive), the difference between the compensating and equivalent measures of welfare change will lie in the range identified in the first two cases. The result of case (a) is predicated upon costless postchange adjustments in holdings while that of case (b) is predicated on the impossibility of postchange adjustments.[3] Positive transactions costs impede adjustments in holdings, whereas either prohibitive transactions costs or indivisibility is sufficient to preclude such adjustments.

In pragmatic benefit-cost analysis terms, these findings may be interpreted as follows. When one is dealing with changes in any one individual's holdings of a divisible good which is traded in reasonably competitive and reasonably low-cost markets, the error inherent in using price multiplied by the quantity change is most likely overshadowed by inaccuracies in the estimation of, for example, the technical productivity coefficients which underlie the quantity change projections.[4] For changes in the quantities of indivisible or lumpy goods where the price flexibility of income falls in the typical range and the proportion of total budget spent on the good is small, the bounds within which the Hicksian compensating and equivalent measures must fall can be rigorously calculated from estimates of Marshallian consumer's surplus; and these bounds will be quite narrow. Thus, the theoretically correct measure can be obtained from the estimates of more readily available value indicators.

However, the analyst is on occasion confronted with the task of evaluating lumpy changes in quantities of specific goods, services, or amenities, where one or more of the following holds: the change is large; the good is highly valued; and the price flexibility of income for the good is high and rises with income. In such cases,[5] our findings offer less comfort. The compensating measure of welfare loss (gain) is likely to be much larger (smaller) in absolute value than the equivalent measure, since the latter (former) is strictly limited by individual budget constraints while the former (latter) is not. While the rigorous bounds derived below remain valid, the information needed to estimate the correct (i.e., compensating, if the welfare criterion is the potential Pareto improvement)[6] measure from other measures more readily available to the analyst is less likely to be obtainable. Thus, in such cases, there may be no good substitute for methods capable of accurately estimating the compensating measure of welfare loss.

In short, the cases that competent benefit-cost analysts consider relatively hazard free are, in fact, relatively hazard free. On the other hand, there are cases (discussed in the preceding paragraph) which, as some benefit-cost analysts have suspected, pose the threat of significant error from the use of Marshallian consumer's surplus and, at the same time, present empirical difficulties to those who seek to estimate the correct measure of welfare change from other more readily available measures. Our contribution will be to demonstrate in a rigorous manner that the intuitions of many benefit-cost analysts are in fact valid, and to show that, in those cases which we have indicated are relatively simple, consumer's surplus may be used "without" apology.

---

[3] The impossibility of postchange adjustments along a quantity continuum, which may be called indivisibility or lumpiness from the perspective of the individual, may arise from any of several sources: 1) the Samuelsonian public good and congestible public good situations in which the individual consumer cannot continuously adjust his holdings of the good independently of all other individuals; 2) indivisibility in production, such that the good may be produced in one or more discrete quantities and the concept of a parametric unit price is meaningless (the argument of David Bradford is helpful); 3) nonexclusiveness, such that exchange in order to effect postchange adjustment is infeasible; 4) prohibitive transactions costs; and 5) institutional restrictions upon quantity per capita, for example, in the case of hunting where regulations such as bag limitations and hunting seasons prohibit quantity adjustments following changes in these kinds of regulations.

[4] This claim is analogous to Willig's claim (p. 589) that the errors from using Marshallian consumer's surplus will often be overshadowed by errors in estimation of the demand curve.

[5] Proposed projects or programs which have the potential to significantly modify unique environments, endangered species, threatened cultures, or the life and health expectancies of human beings may, under certain circumstances, provide examples.

[6] See Mishan.

I. A Diagrammatic Exposition

Our conclusions are perhaps most readily grasped, at the intuitive level, from a diagrammatic exposition. Let us consider a normal good x, which, depending on the program alternatives chosen, may be provided in two different quantities, Q' and Q'', where Q'' is greater and *ceteris paribus* preferred. The traditional pragmatic measures of value of these two bundles of goods are willingness to pay (*WTP*) and willingness to accept (*WTA*). In a market exchange situation, these correspond, respectively, to the buyer's best offer and the seller's reservation price; in a nonmarket situation, they correspond to willingness-to-pay and willingness-to-accept compensation.

If the program alternative under evaluation would reduce the quantity of x from Q'' to Q', the compensating measure of welfare loss is WTA^C, the compensation which would keep the loser at his initial welfare level; and the equivalent measure is WTP^E, the loser's willingness to pay to avoid the quantity reduction from Q'' to Q' which, if paid, would place the loser at his subsequent welfare level. If the proposed program would increase the quantity of x from Q' to Q'', the compensating measure of welfare gain is WTP^C which, if paid, would keep the gainer at his initial welfare level; and the equivalent measure is WTA^E, the compensation which would be needed to bring the potential gainer to his subsequent welfare level in the event that the proposed program is not implemented. At this point, let us emphasize that *WTA* is not always the compensating measure and *WTP* is not always the equivalent measure.

Assume x is a perfectly divisible good, traded in infinitely large markets at zero transactions costs at the unit price p. Consider a proposed program which if implemented would reduce an affected individual's holdings of x from Q'' to Q' while leaving his holdings of Y, a numeraire or a composite of "all other goods," at $\bar{Y}$. In Figure 1, the program would bump the individual from point E to B, lowering his welfare level from I'' to I'. However, there is no good reason for the individual to remain at B. Instead, frictionless markets will permit him to trade along his new budget line until he reaches D and achieves the welfare level I^*. Given this frictionless adjustment, his WTP^E is EF, which is equal to $Y''Y'$ while his WTA^C is BC, which is equal to $Y'Y''$. Thus, WTP^E is equal to WTA^C and both are equal to $p(Q''-Q')$.

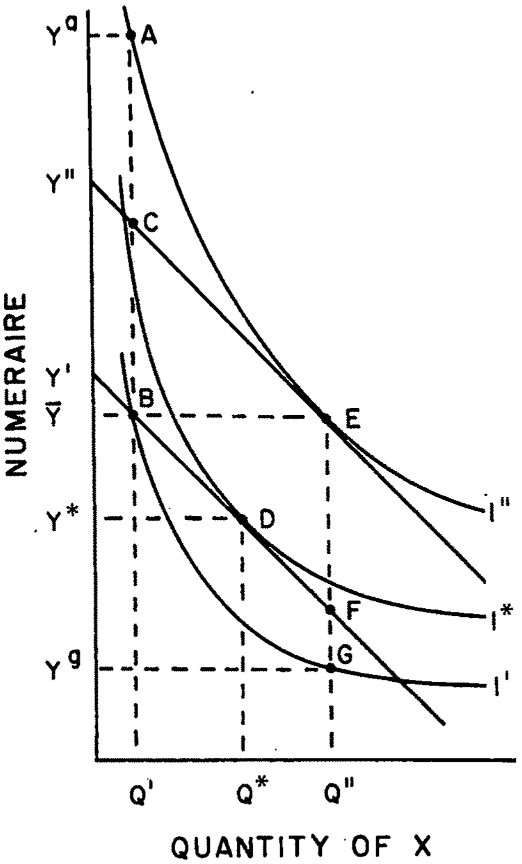


FIGURE 1. THE WELFARE IMPACT OF A CHANGE IN THE QUANTITY OF A GOOD X FROM Q'' to Q'

This conclusion can be grasped intuitively by considering that a good traded in infinitely large markets at a constant unit price with zero transactions costs has all the important characteristics of money (i.e., currency). Thus, the well-known result, that *WTA* compensation and permit imposition of a lump sum tax is equal to *WTP* to avoid the same tax and both are equal to the tax itself, applies to this quite restrictive case.

Now, assume that x is a lumpy good and can be held only in the amounts Q'' and Q'. Observe immediately that in the case of indivisible or lumpy goods, since intermediate adjustments in commodity holdings are not permissible, the Hicksian compensating and equivalent measures in commodity space are analogous to the Hicksian surpluses, not the variations, defined over price changes.

In this case and returning to Figure 1, the price lines become meaningless. WTP^E is EG which is equal to $Y^g\bar{Y}$, and WTA^C is BA, which is equal to $\bar{Y}Y^a$, and larger in absolute value than WTP^E. Note that this finding contrasts vividly with that of Mäler (pp. 131–140), but that Mäler's analysis is valid only if one assumes, trivially, that the individual is indifferent as to whether he holds the indivisible good or not, when its price is zero.

II. Bounds on the Difference between Compensating and Equivalent Measures, Defined in Commodity Space

Now, we derive rigorously the bounds on $WTA - WTP$ for indivisible or lumpy goods, using a formulation which proves, as a special case, that for perfectly divisible goods traded in frictionless markets, $WTA = WTP$. For the convenience of readers and to permit an otherwise unattainable brevity, our analysis at this point follows Willig. The necessary departures from Willig's treatment are explicitly highlighted and explained, while some intermediate steps analogous to those used by Willig are omitted.

Postulating the existence of the utility function $U[P(Q,Y)]$, implicit in prices p, and defined over the numeraire Y, and quantities Q of the good x, we define the following indirect utility function

$$z(Q,Y) = U[P(Q,Y)] \tag{1}$$

Consider an individual with $\bar{Y}$ units of the numeraire (i.e., the value of "all other goods and endowments"), who is confronted with a change in his bundle of the good x from Q'' to Q', where Q'' represents a greater quantity than Q'. His WTP^E to avoid a change from Q'' to Q' is defined by

$$z(Q',\bar{Y}) = z(Q'',\bar{Y} - WTP) \tag{2}$$

while his WTA^C to accept a change from Q'' to Q' is defined by

$$z(Q',\bar{Y} + WTA) = z(Q'',\bar{Y}) \tag{3}$$

Using the income compensation function, $\mu(Q|Q'',\bar{Y})$ which represents the least amount of income the individual would require with quantity Q to achieve the same utility level enjoyed with quantity Q'' and $\bar{Y}$, the individual's initial income is

$$\mu(Q''|Q'',\bar{Y}) = \bar{Y} \tag{4}$$

The *WTP* and *WTA* measures defined in (2) and (3) can be expressed

$$WTP = \mu(Q'|Q',\bar{Y}) - \mu(Q''|Q',\bar{Y}) \tag{5}$$

$$WTA = \mu(Q'|Q'',\bar{Y}) - \mu(Q''|Q'',\bar{Y}) \tag{6}$$

Now the following partial differential equation may be derived (since, although we later define x as a lumpy good, preferences are continuous over Q):

$$\frac{\partial \mu(Q|Q'',Y)}{\partial Q} = P[Q,\mu(Q|Q'',\bar{Y})] \tag{7}$$

Equation (7) may be treated as a system of equations when x is a vector of goods.

Considering only a change in the quantity of a single good and using the fundamental theorem of calculus,

$$WTP = \int_{Q'}^{Q''} P[Q,\mu(Q|Q',\bar{Y})]\,dQ \tag{8}$$

$$WTA = \int_{Q'}^{Q''} P[Q,\mu(Q|Q'',\bar{Y})]\,dQ \tag{9}$$

These equations express *WTP* and *WTA* as areas under compensated demand curves. The price axis consists of a range of implicit prices which corresponds to the domain of quantities being considered.

To arrive at bounds for the difference between the Marshallian measure of welfare change, and *WTP* and *WTA*, respectively, for a quantity change which permits no subsequent adjustments in the quantity of x, we assume constant price flexibility of income, ζ, for x such that

$$(\partial P(Q,Y)/\partial Y)(Y/P(Q,Y)) \equiv \zeta$$

and follow a procedure similar to Willig's to derive the income compensation function, which is

$$(10)\quad \mu(Q'|Q'',\bar{Y}) = \bar{Y}\left[1+\frac{(1-\zeta)}{\bar{Y}}\int_{Q'}^{Q''}P(Q,\bar{Y})dQ\right]^{1/(1-\zeta)}$$

The expression $\int_{Q'}^{Q''} P(Q,\bar{Y})dQ$ can be interpreted as the area under an implicit Marshallian demand curve for the good x. Denoting this expression by M, it is apparent that when $\zeta=0$ the welfare change measures will be equal to M.

Again following procedures similar to those used by Willig, we obtain

$$(11)\quad \frac{M-WTP}{M} \simeq \frac{\zeta M}{2\bar{Y}}$$

$$(12)\quad \frac{WTA-M}{M} \simeq \frac{\zeta M}{2\bar{Y}}$$

Thus,

$$(13)\quad WTA-WTP \simeq \frac{\zeta M^2}{\bar{Y}}$$

Dropping the assumption of constant price flexibility of income (see Willig, who dropped the assumption of constant income elasticity of demand) the following approximate bounds are appropriate when $\zeta M/2\bar{Y}$ is small (say, $\leqslant .05$), and are obtained for *WTP* and *WTA* for changes in the bundle of goods:

$$(14)\quad \frac{\xi^L M}{2\bar{Y}} \leqslant \frac{M-WTP}{M} \leqslant \frac{\xi^u M}{2\bar{Y}}$$

$$(15)\quad \frac{\xi^L M}{2\bar{Y}} \leqslant \frac{WTA-M}{M} \leqslant \frac{\xi^u M}{2\bar{Y}}$$

where ξ^L and ξ^u are the lower and upper bounds, respectively, on price flexibility of income for the good x. Rigorous bounds, applicable for all values of $\zeta M/2\bar{Y}$, are

$$(16)\quad \frac{\left[1-(1-\xi^L)\frac{M}{\bar{Y}}\right]^{1/(1-\xi^L)}-1+\frac{M}{\bar{Y}}}{M/\bar{Y}} \leqslant \frac{M-WTP}{M} \leqslant \frac{\left[1-(1-\xi^u)\frac{M}{\bar{Y}}\right]^{1/(1-\xi^u)}-1+\frac{M}{\bar{Y}}}{M/\bar{Y}}$$

$$(17)\quad \frac{\left[1+(1-\xi^L)\frac{M}{\bar{Y}}\right]^{1/(1-\xi^L)}-1-\frac{M}{\bar{Y}}}{M/\bar{Y}} \leqslant \frac{WTA-M}{M} \leqslant \frac{\left[1+(1-\xi^u)\frac{M}{\bar{Y}}\right]^{1/(1-\xi^L)}-1-\frac{M}{\bar{Y}}}{M/\bar{Y}}$$

The critical differences between these bounds for quantity changes and Willig's bounds for price changes are M and ζ in our bounds, and A and η in Willig's. The term A is Marshallian consumer's surplus measured as the area under a demand curve between two alternative prices for x, while M is the area under a demand curve between two quantities of x. Our ζ is the price flexibility of income for x, while Willig's η is the income elasticity of demand for x. An additional difference is that Willig's m is total income, while our $\bar{Y}$ is the amount of the numeraire held by the consumer.

Returning now to the case of a perfectly divisible good traded in frictionless markets, the expressions $P[Q,\mu(Q|Q',\bar{Y})]dQ$ (in equation (8)) and $P[Q,\mu(Q|Q'',\bar{Y})]dQ$ (in equation (9)) can both be replaced with the market price p of the good x. Thus

$$WTP=\int_{Q'}^{Q''} p\,dQ=WTA \tag{18}$$

and it is shown that WTP and WTA are equal in value and both equal to $p(Q''-Q')$.

For completeness, it is appropriate to note that for goods which are exchangeable and divisible, the difference between WTA and WTP becomes nonzero when transactions costs become positive and grows as transactions costs increase. It will, however, reach an upper limit at the point where transactions costs are just sufficiently great to preclude postchange quantity adjustments and will at that point be identical to the difference observed in the case of lumpy goods.[7]

III. Empirical Estimates of the Bounds

Willig presents empirical estimates of the parameters necessary to calculate the bounds on the difference between compensating, equivalent, and Marshallian consumer's surplus measures of the welfare effects of price changes (his Table 1). That table can be adapted to permit calculation of these bounds for measures of the welfare effects of changes in the bundle of goods. If one substitutes $M/\bar{Y}$ for Willig's a in each column heading and ζ for η in each row title, the first entry in each cell of Willig's table becomes an estimate of $\zeta M/2\bar{Y}$ and may be used in empirical application of equations (11)–(13).

Making these adaptations and comparing the first entry in each cell with the second and third entries, it becomes clear that the rule of thumb for approximating WTA, M, or WTP from empirical estimates of any one of them (i.e., equations (11)–(13)) generates quite accurate results when $M/\bar{Y}$ is small, even in cases where ζ is substantially greater than 1. However, as $M/\bar{Y}$ becomes large, the error from using the rule of thumb grows; it becomes very large when both $M/\bar{Y}$ and ζ are large. In the case of a good which is indivisible and highly valued (i.e., $M/\bar{Y}$ is large) and for which the price flexibility of income rises with income, M provides only poor estimates of WTA and WTP, and the rule of thumb (equations (11)–(13)) is unreliable. The rigorous bounds, however, remain valid.

Some comment on ζ, the price flexibility of income, is appropriate. For the case of changes in the quantity of a good for which postchange quantity adjustments are impossible, ζ is the appropriate concept of income elasticity. The relevent concept is not η, which addresses the question of how changes in income affect the quantity of the good purchased at some parametric unit price, but ζ, which addresses the question of how changes in income affect the amount the consumer would spend to enjoy a given unit of the good.[8] For normal goods, $\zeta>0$. For an especially treasured good, ζ may substantially exceed 1.

---

[7] In order to demonstrate this diagrammatically, it is possible to modify Figure 1 by introducing transactions costs, in the manner of Jürg Neihans, via a kinked budget constraint which touches the zero transactions costs budget constraint at the point denoting the individual's current holdings of x and moves progressively further away from it, always on the low side, as the quantity adjustment becomes larger. In a working paper available upon request from the authors, this diagrammatic analysis is provided, along with a method for approximating, subject to some restrictive assumptions, the bounds on the difference between WTA and WTP when transactions costs are positive but not sufficiently large to preclude any quantity adjustment. For example, if $M/\bar{Y}=0.02$, the income elasticity of demand for the good x is 0.9 and transactions costs amount to 10 percent of p, the difference between WTA and WTP will be less than 0.36 percent of $p(Q''-Q')$.

[8] The relationship of price flexibility of income, $\zeta=(\partial P(Q,Y)/\partial Y)(Y/P(Q,Y))$, to income elasticity of total value, $(\partial V/\partial Y)\cdot(Y/V)$, is obvious when one considers that $V=\int_{Q'}^{Q''}P(Q,Y)dQ$.

REFERENCES

D. F. Bradford, "Benefit-Cost Analysis and Demand Curves for Public Goods," *Kyklos*, No. 4, 1970, *23*, 775–91.

C. J. Cicchetti, A. C. Fisher, and V. K. Smith, "An Econometric Evaluation of a Generalized Consumer's Surplus Measure: The Mineral King Controversy," *Econometrica*, Nov. 1976, *44*, 1259–76.

Karl-Göran Mäler, *Environmental Economics: A Theoretical Inquiry*, Baltimore 1974.

E. J. Mishan, *Cost-Benefit Analysis*, London 1971.

J. Neihans, "Money and Barter in General Equilibrium with Transactions Costs," *Amer. Econ. Rev.*, Dec. 1971, *61*, 773–83.

R. D. Willig, "Consumer's Surplus Without Apology," *Amer. Econ. Rev.*, Sept. 1976, *66*, 587–97.

On the Effect of Rate of Return Regulation under Uncertainty

By SATYA P. DAS*

Recently some attempts have been made to incorporate uncertainty in the theory of a regulated firm (see, for example, Stylianos Perrakis, and Yoram Peles and Jerome Stein). These authors argue that under uncertainty the Averch-Johnson (hereafter A-J) effect in general does not hold. They base their argument on a test of one of the implications for a traditional regulated firm (under certainty), which runs as follows:

PROPOSITION 1: *A firm under regulation employs a higher amount of capital than it would if it was not regulated.*

This proposition, though correct, can certainly be challenged as the one on which the case for regulation should stand or fall. This is because Proposition 1 is merely a positive statement which does not have any normative content. Rather the most crucial implication for a regulated firm which does have an obvious but important normative content, is as follows:

PROPOSITION 2: *The capital-labor ratio chosen by a profit-maximizing regulated firm will be higher than that consistent with cost minimization for the output it selects.*

This is a very important proposition from the point of view of policy, for it implies that regulation leads to inefficiency in the factor market, while originally intended to alleviate imperfections in the product market. Harvey Averch and Leland Johnson themselves regarded Proposition 2 as one of their fundamental results when they wrote, "The firm does not equate marginal rates of factor substitution to the ratio of factor costs; therefore the firm operates inefficiently in the sense that (social) cost is not minimized at the output it selects" (p. 1052). And it should be noted that Proposition 2 *does not* necessarily follow from Proposition 1. One can easily imagine that some of the implications for a regulated firm under certainty will carry over to the presence of uncertainty and some will not. But it is always worthwhile and interesting to determine whether or not the most crucial implication for a regulated firm extends to the case of uncertainty. In this spirit, the objective of the present paper is to test Proposition 2 under uncertainty rather than Proposition 1 which has been examined by previous authors. The conclusion is affirmative, that is, Proposition 2 (the A-J thesis) does hold under uncertainty. This is proved in two different models of uncertainty: a quantity-setting model almost identical to the one used by Peles and Stein, and a price-quantity-setting model of Edwin Mills' type. These are described in Sections I and II, respectively.

Moreover, compared to a risk-neutral firm considered by Peles and Stein, I have introduced a non-linear utility function for profit. *Surprisingly, my conclusion does not depend on the nature of the utility function or, in other words, on the behavior towards risk such as risk aversion, risk neutrality, or risk preference.* This illustrates how robust the A-J thesis is.

I. A Quantity-Setting Firm

In Peles and Stein's paper, the scale of plant or the level of the capital stock K is the only *ex ante* variable, whereas the price of the product p, and the amount of labor employed L, are *ex post* variables. This is somewhat at variance with the existing literature in which the decision on optimal amounts of K and L is taken simultaneously.[1] In this paper such a simultaneous decision is assumed in order to remain as close to the traditional model as possible. In other words, K and L are *ex ante* variables and (as in Peles and Stein's paper) p is the *ex post* variable.

---

*University of Wisconsin-Milwaukee.

[1] See, for example, Averch and Johnson, William Baumol and Alvin Klevorick, and Elizabeth Bailey.

Let $q=h(p,u)$ be the demand function where q is the quantity demanded and u, the random term. Inverting this function we get $p=p(q,u)=p(f(K,L),u)$ where f is the production function. Then the total revenue R is a function of K, L, and u, i.e., $R=R(K,L,u)$. Without loss of generality, let us assume that $R_u>0$.[2] Quasi rents Z are given by the total revenue minus the labor costs, i.e., $Z=R(K,L,u)-wL$, where w is the given price of labor. Let s denote the fair rate of return on capital, and let us define $u=u_0$ such that $Z=sK$. Then it is clear that the quasi rents are given by $R-wL$ only when $u\leqslant u_0$. If $u>u_0$, they are determined by adjusting the price level. Then the firm's profit is given by

$$\pi_1=R(K,L,u)-wL-rK \qquad \text{for } \alpha\leqslant u\leqslant u_0$$
$$\pi_2=(s-r)K \qquad \text{for } u_0<u\leqslant \beta$$

where α and β are the lower and upper limits of u, and r is the given unit cost of capital.

This model is exactly the same as Peles and Stein's except that labor is an *ex ante* variable. Let us further assume that the von Neumann-Morgenstern axioms are satisfied so that the firm's objective is to maximize the expected utility of profit, i.e.,

$$\operatorname{Max} G(K,L)$$
$$\equiv\int_{\alpha}^{u_0} U(\pi_1)\phi(u)\,du+U(\pi_2)\int_{u_0}^{\beta}\phi(u)\,du$$

where $\phi(\cdot)$ is the density function of u. The first-order conditions are given by[3]

$$(1)\quad G_K\equiv\int_{\alpha}^{u_0} U'(\pi_1)(R_K-r)\phi(u)\,du + U'(\pi_2)(s-r)\int_{u_0}^{\beta}\phi(u)\,du=0$$

$$(2)\quad G_L\equiv\int_{\alpha}^{u_0} U'(\pi_1)(R_L-w)\phi(u)\,du=0$$

The functions R_K and R_L can be written as $R_K=f_K(p+fp_q); R_L=f_L(p+fp_q)$. Using these, (1) and (2) can be written as

$$(1')\quad f_K\int_{\alpha}^{u_0} U'(\pi_1)(p+fp_q)\phi(u)\,du = r\int_{\alpha}^{u_0} U'(\pi_1)\phi(u)\,du - (s-r)\,U'(\pi_2)\int_{u_0}^{\beta}\phi(u)\,du$$

$$(2')\quad f_L\int_{\alpha}^{u_0} U'(\pi_1)(p+fp_q)\phi(u)\,du = w\int_{\alpha}^{u_0} U'(\pi_1)\phi(u)\,du$$

Dividing (1′) by (2′), we get

$$\frac{f_K}{f_L}=\frac{r}{w}-\frac{(s-r)\,U'(\pi_2)\int_{u_0}^{\beta}\phi(u)\,du}{w\int_{\alpha}^{u_0} U'(\pi_1)\phi(u)\,du}$$

It is usually assumed that the fair rate of return is higher than the unit cost of capital, i.e., $s>r$. Thus we have $f_K/F_L<r/w$. This implies, from the standard assumption of a quasi-concave production function, that the optimal K/L ratio is higher than that at the cost-minimization point.

II. A Price-Quantity-Setting Firm

In this model, it will be assumed that the amounts of capital and labor as well as the price level are the *ex ante* variables. In other words, it is a case of a price-quantity-setting firm of Mills' type.[4] This model also keeps the analysis parallel to the standard A-J model insofar as the decisions on optimal K and L are taken at the same time. However a fundamental difference between this model and the one discussed in the previous section is that here the quantity produced is in general different from the actual quantity demanded. This implication, which might appear restrictive specifically for a regulated firm, could be justified in two ways: (a) In the case of public utilities, the common example of a regulated firm, excess demand sometimes does exist, for example, the inability of the telephone company to take some long distance calls around Christmas. Excess supply can of course be easily justified by the existence of excess capacity. (b) Even though most regulated industries today are utilities or services, from the theoretical point of view, the case of a regulated firm producing a commodity (as distinguished from a service) is also worth examining. There, the existence of excess supply or excess demand is very plausible.

---

[2] Throughout this paper, subscripts will denote the partials.

[3] Note that u_0 is a function of K and L. But the terms involving the partials of u_0 cancel out in (1) and (2). It may be pointed out here that the terms $\phi_1(K)$ and $\phi_2(K)$ in Peles and Stein's paper, which are analogous to the partials of u_0 in this paper, are in fact equal to zero. Thus there is no need to ignore $\phi_1(K)$ and $\phi_2(K)$ as Peles and Stein do.

[4] A discussion of such a model can be found in Hayne Leland also.

As in the previous model, let the demand function be $q=h(p,u)$. Without loss of generality, let h_u be positive. Let X denote the output, and q^s, the quantity sold. Then the firm's profit π is given by $pq^s-wL-rK$. By analogy with u_0 in the earlier model, define H such that $q=h(p,H)=X=f(K,L)$. Thus if $u \geqslant H$, then $q \geqslant X$ and $q^s=X$; and if $u \leqslant H$, we have $q \leqslant X$ and $q^s=q$. Thus profit will be

$$\pi_1=ph(p,u)-wL-rK \qquad \text{for } \alpha \leqslant u \leqslant H$$

$$\pi_2=pf(K,L)-wL-rK \qquad \text{for } H \leqslant u \leqslant \beta$$

and hence the expected utility of profit is given by

$$(3) \quad \int_\alpha^H U(\pi_1)\phi(u)\,du+U(\pi_2)\int_H^\beta \phi(u)\,du$$

The rate of return constraint in this model is given by $pq^s-wL \leqslant sK$. But as the upper limit of the quasi rent, given p,K, and L, is $pX-wL$, the rate of return constraint is equivalent to[5]

$$(4) \qquad pf(K,L)-wL \leqslant sK$$

Thus the objective of the firm is to maximize (3) subject to (4), where K, L, and p are the control variables.

There is one more complication to address before deriving the results formally. A classical argument against regulation is that it motivates the firm to employ some redundant capital which is not used in production but purchased simply to raise (or pad) the rate base in the eye of the regulator. However it has been shown (see Edward Zajac or Bailey, for example) that such "rate-base padding" does not help; or, in other words, the optimal amount of redundant capital is zero. Similarly one can conceive of some redundant labor, whose optimal amount also turns out to be zero (see Bailey, ch. 5). Following Bailey, let us, in a general way, introduce variables K^* and L^* which stand for redundant capital and redundant labor, respectively. As in Bailey's text, it considerably simplifies the algebra here.

Now with the additional variables K^* and L^*, the objective of the firm can be written as

$$\begin{aligned}\operatorname*{Max}_{K,L,p,K^*,L^*} \int_\alpha^H U[&ph(pu)-w(L+L^*)\\ &-r(K+K^*)]\phi(u)\,du\\ +U[&pf(K,L)-w(L+L^*)\\ &-r(K+K^*)]\int_H^\beta \phi(u)\,du\end{aligned}$$

subject to

$$pf(K,L)-w(L+L^*) \leqslant s(K+K^*)$$

Let M and λ stand for the Lagrangian and the Lagrangian multiplier of this constrained maximization. Then, assuming positive solutions to K, L and p, the first-order Kuhn-Tucker conditions are given by[6]

$$\begin{aligned}(5) \quad \partial M/\partial K=pf_K U'(\pi_2)&\int_H^\beta \phi(u)\,du\\ &-rE[U'(\pi)]+\lambda(s-pf_K)\\ =0&\end{aligned}$$

---

[5] In fact, assuming that the regulatory constraint is effective, it becomes an equality; but the ineqality form helps in determining the range of the Lagrangian multiplier of this constrained maximization.

[6] Note that H is a function of p, K, and L; but as in Section I, the terms involving its partials cancel out in the first-order conditions.

$$\partial M/\partial L = pf_L U'(\pi_2)\int_H^\beta \phi(u)\,du - wE[U'(\pi)] + \lambda(w - pf_L) = 0 \tag{6}$$

$$\partial M/\partial p = \int_\alpha^H U'(\pi_1)(h + ph_p)\phi(u)\,du + U'(\pi_2)\cdot f\cdot \int_H^\beta \phi(u)\,du - \lambda f = 0 \tag{7}$$

$$\partial M/\partial K^* = rE[U'(\pi)] + \lambda s \leqslant 0 \tag{8}$$

$$(\partial M/\partial K^*)K^* = 0 \tag{9}$$

$$\partial M/\partial L^* = -wE[U'(\pi)] + \lambda w \leqslant 0 \tag{10}$$

$$(\partial M/\partial L^*)L^* = 0 \tag{11}$$

$$\partial M/\partial \lambda = w(L + L^*) + s(K + K^*) - pf(K,L) \geqslant 0 \tag{12}$$

$$(\partial M/\partial \lambda)\lambda = 0 \tag{13}$$

$$\lambda \geqslant 0 \tag{14}$$

I will now prove two lemmas which will lead us to a proof of Proposition 2—my final objective.

LEMMA 1 (*Lagrangian Multiplier*): *If the rate of return constraint is effective, then* $0 < \lambda < E[U'(\pi)]$.

Needless to say, this lemma has a perfect counterpart in the standard A-J model, which is $0 < \lambda < 1$.[7]

PROOF:

When the constraint is effective, $\lambda > 0$, for otherwise the first-order conditions reduce to those of an unregulated monopoly. Moreover, expression (10) on rearrangement implies $\lambda \leqslant E[U'(\pi)]$. Thus it remains to show that $\lambda \neq E[U'(\pi)]$. Suppose $\lambda = E[U'(\pi)]$. Then expression (6) reduces to

$$pf_L U'(\pi_2)\int_H^\beta \phi(u)\,du - \lambda w + \lambda w - \lambda pf_L = 0$$

implying $\lambda = U'(\pi_2)\int_H^\beta \phi(u)du$. Substituting this in expression (5), we get $\lambda pf_K - \lambda r + \lambda s - \lambda pf_K = 0$, which implies that $s = r$, a contradiction to the usual assumption that $s > r$. Hence the proof.

LEMMA 2 (*Rate-Base Padding*): *The optimal* K^* *and* L^* *are zero.*[8]

PROOF:

From Lemma 1, $\lambda < E[U'(\pi)]$ which implies from expression (10) that $\partial M/\partial L^* < 0$, which in turn from expression (11) implies that $L^* = 0$. Now, if $K^* \neq 0$, then from expression (9), $\partial M/\partial K^* = 0$; so that $\lambda s = rE[U'(\pi)]$. Substituting this in expression (5), we get $\lambda = U'(\pi_2)\int_H^\beta \phi(u)du$. But expression (6) can be written as

$$pf_L\left[U'(\pi_2)\int_H^\beta \phi(u)\,du - \lambda\right] = [E\{U'(\pi)\} - \lambda] \tag{15}$$

which implies through Lemma 1 that $\lambda < U'(\pi_2)\int_H^\beta \phi(U)du$. Thus $K^* \neq 0$ leads to a contradiction. Hence $K^* = 0$.

Lemma 2 ensures that the solutions for K, L, and p given by expressions (5)–(14) are also the solutions to the maximization of (3) subject to (4).

We may now prove the central theorem, i.e., $f_K/f_L < r/w$, which immediately implies Proposition 2.

THEOREM: *At the optimal level,* $f_K/f_L < r/w$.

PROOF:

Let us rewrite expression (5) as

$$pf_K\left[U'(\pi_2)\int_H^\beta \phi(u)\,du - \lambda\right] = r[E\{U'(\pi)\} - \lambda] - \lambda(s - r) \tag{16}$$

---

[7] See Bailey, p. 74, Proposition 5.1.

[8] This result also holds under certainty. See, for example, Bailey, p. 75, Proposition 5.2.

Dividing (16) by (15) we get

$$\frac{f_K}{f_L}=\frac{r}{w}-\frac{\lambda(s-r)}{w[E\{U'(\pi)\}-\lambda]}$$

where s is greater than r by assumption, and from Lemma 1, $E[U'(\tau)]-\lambda$ is positive. Hence $f_K/f_L<r/w$.

REFERENCES

H. Averch and L. L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

Elizabeth E. Bailey, *Economic Theory of Regulatory Constraint*, Lexington 1973.

W. J. Baumol and A. K. Klevorick, "Input Choices and Rate of Return Regulation: An Overview of the Discussion," *Bell J. Econ.*, Autumn 1970, *1*, 162–90.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

E. J. Mills, "Uncertainty and Price Theory," *Quart. J. Econ.*, Feb. 1959, *73*, 116–29.

Y. C. Peles and J. L. Stein, "The Effect of Rate of Return Regulation is Highly Sensitive to the Nature of Uncertainty," *Amer. Econ. Rev.*, June 1976, *66*, 278–89.

S. Perrakis, "Rate of Return Regulation of a Monopoly Firm with Random Demand," *Int. Econ. Rev.*, Feb. 1976, *17*, 149–62.

______, "On the Regulated Price-Setting Monopoly with a Random Demand Curve," *Amer. Econ. Rev.*, June 1976, *66*, 410–16.

E. E. Zajac, "A Geometric Treatment of the Averch-Johnson Behavior of the Firm Model," *Amer. Econ. Rev.*, Mar. 1970, *60*, 117–25.

Limits on Public Provision of Private Goods

By ROBERT J. STAAF AND E. G. WEST*

In a model that incorporates variable benefit shares into a median voter analysis, Arthur Denzau and Robert Mackay (hereafter, D-M) extend the range of goods that may be publicly provided to include private commodities where price discrimination is feasible.[1] One alleged result is that if the tax price p_i^q to a majority of voters is less than the market price p of a privately tradeable good, where price discrimination is *not* possible, then an unbounded increase in the public provision of the private good will occur since the majority can sell the excess beyond their own market quantity demanded, making a per unit profit of $(p - p_i^q)$. Public provision of the private good under these circumstances is therefore improbable.

Our comment concentrates on this result of the D-M analysis. Their central contributions of benefit shares and general analysis of price discrimination are not at issue here. Their finding does not consider the effects on the price of the privately tradeable good as the market equilibrium quantity is exceeded. We show that public provision of private goods may be undertaken to transfer income indirectly from the minority to the majority, and that such income transfers will be maximized when the equilibrium quantity publicly provided is identical to the market equilibrium quantity. Hence it will never be in the interests of a majority coalition to extend public provision beyond a market equilibrium quantity. Under certain conditions, public provision of a private good with subsequent market resale is basically a change in the property right structure and may be examined in the context of Coase's neutrality theorem.[2]

I. The Basic Model

Denzau and Mackay state their case for the unlikely public provision of private goods as follows:

> Combining these results leads to the conclusion that for the *public provision* of a purely *private good* to continue in the presence of a private, competitive market, allowing for private purchases and resales, a majority of voters must have *tax prices* p_i^q of *exactly* p. In short, continued public provision of a freely exchangeable private commodity requires either: 1) outlawing private exchange of the good; 2) a special earmarked tax with tax shares set equal to benefit shares for a majority of voters; or 3) the setting of benefit shares so that they equal the marginal tax shares implicit in the given finance system. Since the occurrence of either of the last two conditions seems improbable, and the criminalization of private markets generates many other problems, it appears unlikely that easily exchangeable goods such as bread could ever be publicly provided in a participatory democracy.
>
> [pp. 74–75, emphasis added]

These conditions are indeed unlikely. The following is a median voter model based on tax shares not equal to benefit shares and the existence of a market (legal or illegal) to illustrate a determinate and bounded solution.

---

*Law and Economics Center, University of Miami (on leave from Virginia Polytechnic Institute and State University), and economics department, Carleton University, respectively. We gratefully acknowledge comments by Mackay who suggested the use of a Cobb-Douglas utility function to rectify criticisms of the referee. We also acknowledge helpful comments by James Buchanan and an anonymous referee. This research was supported, in part, by the Ford Foundation and the Law and Economics Center, University of Miami.

[1] The analysis differs from Robert Spann's since his model like D-M assumes a private good where price discrimination is possible.

[2] The usual caveat of ignoring transactions costs is assumed.

1) The private good q is produced in a competitive industry where $p = c$ and marginal cost c is constant.

2) A free exchange market is permitted (or exists) after the collective allocation of the good and transactions costs are zero.

3) Benefit shares are equal such that $\alpha_i = 1/m$ where m is the total number of recipients of the publicly provided private good.

4) Taxes are levied on the basis of a proportional income tax such that the tax share, τ_i is equal to $\tau_i = y_i / \Sigma_j y_j; i,j = 1, \ldots, m$.[3]

5) The distribution of income is symmetrical where y median $= y$ mean.

6) Cobb-Douglas utility functions are assumed.

Given these assumptions the median (income) voter will face a tax price p_i^q that is equal to the market determined price p, since his tax share τ_i equals his benefit share α_i.[4] Individuals with less (greater) than median income will have tax prices, p_i^q, less (greater) than p since their tax share, τ_i, is less (greater) than the average or median tax share. These conditions will be illustrated with a simple three-person model.

Refer to Figure 1. Prior to collectivization, and given a market price of \$1, individual 1 is at equilibrium at point E_1 and demands two units. Individual 2, at equilibrium point E_2, demands four units. Individual 3, at equilibrium point E_3, demands six units. Aggregate supply is thus twelve units. Under the assumptions of a symmetrical income distribution and unitary income elasticity, individual 1 initially has 0.5 of the median (mean) income of individual (2). Individual 3 has 1.5 the income of individual 2. The units of incomes on the vertical ordinate are four, eight, and twelve, respectively.

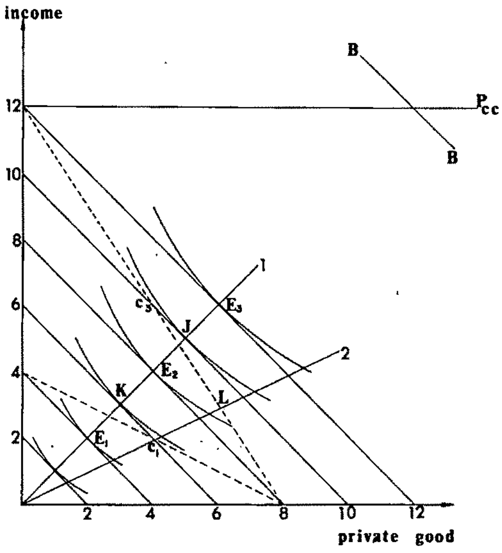


FIGURE 1

Upon collectivization, the median voter, individual 2, will vote for a total collectively provided amount of twelve units. He will receive an equal share of four units at a tax price p_i^q per unit, equal (in this case) to the market price p. His position will be identical to the market outcome. Individual 1's tax share is 0.5 of individual 2's since the tax rate is proportional to income. The (lower) tax price p_i^q of this individual is shown in the dashed opportunity line which, immediately after collectivization, will lead him to vote for four units (point c_1). Individual 3's tax price p_i^q is 1.5 of individual 2's or 1.5 times the market price and will lead him to vote for four units also (point c_3) assuming the absence of private alternatives.

As shown in Figure 1, individuals 1 and 2 have different marginal evaluations at the fourth unit of the private good collectively and uniformly provided to each. Thus there are potential gains from trade between them. In a large number setting, the exchange price will equal the previously determined market price. Therefore individual 1 will trade one unit of the private good to individual 3 in return for \$1 in income. The new, posttrading equilibria are thus located on the ray 1 at J, E_2, and K, bringing about an equality of the marginal rates of substitution.

---

[3] Alternatively, we could assume $\tau_i = 1/m$: (equal tax shares) and $\alpha_i = \mu y_i^\gamma / \Sigma_j y_j^\gamma$ where $\mu = 1$ and $\gamma < 1$: regressive benefit shares (see D-M, p. 72).

[4] That is, $\tau_i = \alpha_i$ for the median voter only such that his income constraint is $y_i \geq z_i + (\tau_i c / \alpha_i) q_i = z_i + p q_i$ where $c = p$ at a market equilibrium quantity.

The aggregate quantity of twelve units supplied and demanded is the same as the previous market-provided quantity. This result is obtained because, after the collective provision of twelve units (with each receiving four units), individuals 1 and 3 move in opposite directions—individual 3, as a buyer, moving to the right on the diagram, and individual 1, as a seller, moving to the left. The only difference between this and the market result is that an income transfer has taken place. Individual 1 at K has received the equivalent of an income increase of two units; individual 3 has experienced an income equivalent decrease of two units. In this new equilibrium position, however, tax shares and price consumption curves will change. This will result in a second sequence of voting that will eventually converge to an identical (equal) income distribution.

Individual 2 in this example is actually indifferent to collective allocation or market allocation. Individual 3 will vote for zero collective action since his tax price is greater than the market price. Individual 1 prefers collective allocation since he earns a per unit profit of $p-p_i^q$ on each unit when the publicly provided quantity is identical to the market equilibrium quantity. Will individual 1 vote for an *unbounded* amount of the private good to be publicly provided as D-M conclude?

II. The Convergence To Income Equality

A. The Open Economy

Our model on its assumptions has already shown a convergence to income equality. This alone illustrates a bounded solution. To go further, assume for a moment that the median individual 2 will remain in the coalition with individual 1 for all proposed quantities in excess of four units each. Consider an extreme example of a proposal for a total public supply of twenty-four units, or eight units each. This supply would exhaust individual and total community income. If we assume an open economy extending beyond this community of three, and the increased supply does not affect the world price assumed equal to $1, then each individual will have a final equilibrium of E_2 where each individual will sell four units of his collective provision of eight units in the world market receiving $4 in income. The end result is an equal distribution of income, but the total public provision of the private good is bounded by the community's resources.

B. The Closed Economy

Now consider the closed economy of these three individuals with a public provision of twenty-four units. All individuals will have an excess supply of the private good with the end result an equal distribution of the private good. Obviously, individuals 1 and 2 find this equilibrium inferior to the market solution and would not vote for public provision. Therefore, the quantity publicly provided will be less in a closed economy than in an open economy.

Consider a collective provision of eighteen units, six units each, in a closed economy. This would place individual 3 before trade at point L in Figure 1. On our assumption the marginal rate of substitution a point L is equal to that at point c_1 on the ray 2. The marginal rate of substitution at point L and c_1 is equal to 0.5 (individual 1's tax price). In order for individual 3 to engage in exchange and demand more than his public provision, the exchange price must be below $0.5. However, individual 1's tax price is $0.5 and individual 2's tax price is $1. Therefore, if they engage in exchange with individual 3 at an exchange price less than $0.5, they will suffer a net loss on the transaction. Since individual 1 would not be willing to lower the price below his tax price of $0.5, there will be no buyers. All three individuals are worse off relative to a collective provision equal to the market equilibrium quantity.

By similar reasoning we can eliminate all collective supply amounts from eighteen units to twelve units. Collective provision in excess of twelve units must entail a reduced exchange price, but individual (2) will not vote for a provision in excess of twelve units because his tax price is equal to $1 which is equal to marginal cost ($p_2^q=c$). Collective provisions that are less than or equal to the market equilibrium are the only possible equilibrium and a public provision equal to the market equilibrium will maximize transfers.

This can be readily seen by referring to the upper right-hand corner of Figure 1. Because of the Cobb-Douglas assumption we can sum the individual price consumption curves to arrive at a community price consumption curve (indicated by the horizontal line P_{cc}) regardless of the type of tax structure assumed (regressive, proportional, or progressive), and regardless of the income effects brought about by the public provision of private goods. The community price consumption curve intersects the production-possibilities curve ($B-B'$) at the market equilibrium quantity of twelve units. Public provision at less than twelve units without private suppliers will mean potential gains from trade are not being exploited and income transfers not maximized. Public provision in excess of twelve units results in a net social and individual welfare loss as previously demonstrated.

III. Other Utility Functions

A. *Symmetrical Income Distribution*

If we now relax the Cobb-Douglas utility function assumption, the analysis becomes more complicated, but the general results are unaffected. The individual income elasticities may differ and attempts to sum the price consumption curves will be dependent upon the tax structure assumed and the resultant redistributional effects of exchange after the public provision. Because of different tax structures and income effects, the community price consumption curve is likely to be radically different from the community price consumption curve under a market solution. That is, the optimal public provision to maximize transfers may differ from the market equilibrium quantity by intersecting the production-possibilities curve to the left or right of a market provision. However, the principle remains that the public provision of a private good will be identical to market provision after incorporating the income effects of different tax prices and the postredistributional effect of exchange.

While individuals with less than the median income will gain from public provision of a private good, assuming a market equilibrium quantity, the median voter (income) receives no benefit from collectivization since his tax price equals the market price. The restrictions of proportional income taxes and a symmetrical income distribution insure that the *average* tax share equals the median tax share. Equal benefit shares insure that the average benefit share will equal the average tax share for the median voter which together imply his tax price equals the market price.[5]

B. *Positively Skewed Income Distribution*

Now assume more realistically that the median income is less than the mean income. Under proportional income taxation, the *median* voter's tax share is less than the mean income voter's tax share.[6] Equal benefit shares implies that the median (income) voter's tax price is less than the market price; i.e., $\alpha_i/\tau_i>1$ or $\tau_i p/\alpha_i=p_i^q<p$. Note that p in both sides of the preceeding equation are identical only when the private good is publicly provided in an amount less than or equal to the market equilibrium quantity. An overprovision will result in an exchange price lower than marginal cost c of production, assuming a competitive industry where $c=p$. Therefore, for our purposes, the equation should be written as $\tau_i c/\alpha_i=p_i^q<p$, where $c=p$ when the public provision is less than or equal to the market equilibrium. This formulation permits us to examine the median voter's tax price of an overprovision where $c>p$. In this case, $\tau_i c/\alpha_i=p_i^q>p$, even though $\alpha_i/\tau_i>1$ as long as $c/p>\alpha_i/\tau_i$.

Stated in words, if the ratio of the marginal cost to price is greater than the ratio of benefit share to tax share, the median voter will pay a tax price greater than the market price and find it advantageous to reduce the publicly supplied amount or opt for the market allocation.[7] Public provision at the market equilibrium quantity, or less, results in the median voter paying a tax price less than the market price generating consumer's surplus and/or a profit on exchange. Therefore, individuals with income less than or equal to the median income who form a majority (voter) coalition will be motivated to vote for a publicly provided quantity equal to or less than the market equilibrium quantity to maximize their transfer.[8]

---

[5] At a publicly provided quantity less than or equal to the market equilibrium quantity which insures $p=c$.

[6] This is true whether a regressive, proportional, or progressive income tax schedule is used.

IV. The Lindahl Solution with Price Discrimination

The result reached here is that, in contrast to Denzau and Mackay's conclusion, the public provision of private goods, where price discrimination is ruled out, is possible. The voting motivation moreover is to redistribute income. The resultant income furthermore will tend to mirror the underlying distribution of property rights in a democratic system of one man-one vote.

As a final note we wish to take issue also with D-M's conclusion that private goods that allow price discrimination not only can be produced publicly, but will involve "stronger" redistribution than the public provision of the private goods (theoretically) would. Taking as an example the case of fire protection, they argue redistribution will take the form of disproportionate shares of the service distributed to taxpayer-beneficiaries, those nearest a fire station, for instance, receiving more (better) services than others. Under Lindahl taxes, the price is equated to the summed marginal rate of substitution (*MRS*) for the public good and each taxpayer faces a tax price equal to his individual *MRS*. Supposing price inelastic demand for fire services, then the Lindahl tax prices will be *lower* for the consumers closer to the fire station. Thus the redistributive effect is strengthened: voters who receive more fire services pay smaller tax bills and enjoy better service.

Our disagreement here is based on the fact that the Lindahl solution operates under a regime of unanimous voting. The location of the fire station is itself a variable that must be voted on. With a Wicksellian system of simultaneous voting and expenditure, no group of taxpayers is going to agree, *ex ante*, that another group receive above average shares of the service without some compensatory additional payment. Under unanimous voting, disproportionate shares can be vetoed in advance by a single voter. We agree that redistribution will take place under majority voting. But then, the price discrimination case is no different from other private goods cases—as argued above.

REFERENCES

A. T. Denzau and R. J. Mackay, "Benefit Shares and Majority Voting," *Amer. Econ. Rev.*, Mar. 1976, *66*, 69–76.

R. M. Spann, "Collective Consumption of Private Goods," *Publ. Choice*, Winter 1974, *20*, 63–82.

---

[7] Note that if the market good *must* be provided collectively, the minority will also prefer a market equilibrium supply since a collective overprovision will mean an overprovision for all even after exchange. Therefore if private market adjustments are prohibited, although the exchange is permitted, the market equilibrium is the only equilibrium that would pass under a rule of unanimity.

[8] The previous assumption of unitary income elasticity is not necessary, as discussed previously.

Limits on Public Provision of Private Goods: Reply and Further Analysis

By ARTHUR T. DENZAU AND ROBERT J. MACKAY*

In a final section to our article in this *Review*, we pointed out problems that might arise in a democratic setting from attempts to provide publicly private goods that are easily exchangeable. In particular, we considered the case in which the political collectivity in question was a small part of a larger, open economy so that the publicly distributed good could be resold to individuals outside the collectivity. We then attempted to deduce the restrictions on the distribution of benefit shares and tax shares that were necessary for an equilibrium to involve public provision of the good at a finite level.

In simplest terms, we argued that if, for example, the town of Blacksburg attempted to provide bread publicly and a majority of its citizen-voters faced tax prices less than the price at which they could export the good to the rest of the economy, then this majority coalition would support unbounded increases in the level of public provision of bread. The pursuit of these "arbitrage profits" by the majority coalition would imply, of course, bankruptcy for the minority coalition. These types of problems, we argued, made it unlikely that easily exchangeable private goods would be publicly provided.[1]

Robert Staaf and E. G. West (hereafter, S-W) have taken issue with this section of our paper on several grounds. First, they argue that even if the collectivity is a small part of an open economy, individuals would not vote for an *unbounded* amount of the private good to be publicly provided since this would "exhaust individual and total community income." They seem to argue that a voter would not support a level of public provision if his implied tax bill exhausted his initial income. This clearly is not true for those voters with tax shares less than their benefit shares. These individuals would be perfectly willing to borrow to finance their tax bills since they earn significant arbitrage profits at the exchange stage. In essence, S-W assume that individuals cannot borrow even when they can so easily arbitrage between the domestic and outside market at a profit. Ruling out borrowing in this context is arbitrary and, of course, it is no surprise that it leads to a bounded solution. As a result, we find this criticism of our conclusion unconvincing.

Staaf and West also assert that in an open-economy setting, public provision of a private good leads, through a sequence of voting, to convergence to income equality. In this case, they are not simply tinkering with our assumptions to generate different results; rather, they are making a serious logical error. They argue that since the outcome resulting from public provision of a private good with exchange within a period is *equivalent* to a certain income increase or decrease, the individual's income in the next period can be treated as if it *actually* were higher or lower by that amount. Alternatively put, they treat the individual's wealth at the exchange stage during one period as the relevant income or wealth variable for the beginning of the next period. This is simply incorrect. The value of an individual's wealth at the exchange stage, denoted w_i, consists of his initial income or endowment y_i, minus his tax bill $\tau_i cx$, plus the value of his share of the distribution of the publicly provided good $\alpha_i px$. This wealth variable does not become the individual's wealth at the beginning of another period and round of voting, but instead constrains the individual's purchases and consumption of private goods at the exchange stage during the current period.[2] Simply because a particular policy is equivalent in its effects in one period to an increase in income does not imply that the individual involved actually has that income to carry over into all future periods. Once this correction is made, their convergence result no longer follows.

---

*University of Arizona and Virginia Polytechnic Institute and State University, respectively. We would like to thank Carolyn Weaver, Robert Tollison, and Geoffrey Brennan for their helpful comments. The research for this paper was supported by the National Science Foundation under Grant SOC76-22438.

[1] We went on to argue that public provision is more likely to involve those types of private commodities, such as services, for which price discrimination is feasible. Thus, the issue is not whether there will be public provision of private goods or not but rather what types of private goods are more likely to be provided publicly.

With S-W's criticisms of our analysis of the open-economy case out of the way, we can turn to the major contribution of their paper. Their attempt to extend our analysis of the public provision of private goods to the closed-economy case raises several interesting issues and, most importantly, makes it quite clear that both of our papers are but subcases of a more general problem. Unfortunately, the particular example S-W have chosen to examine and the manner in which they state their results are potentially misleading, both with respect to the general nature of the problem and the properties of its solution.

Consider, for example, the following from S-W:

> We show that public provision of private goods may be undertaken to transfer income indirectly from the minority to the majority, and that such income transfers will be maximized when the equilibrium quantity publicly provided is identical to the market equilibrium quantity. Hence it will never be in the interests of a majority coalition to extend public provision beyond a market equilibrium quantity. [p. 461]

Such a statement misrepresents both the nature of the problem as being one of simple transfer maximization and the character of the solution as being identical to market provision. In Section I, therefore, we present as briefly as possible the structure of the general problem and its solution. In Section II, the solution to a specific example is presented to illustrate the types of issues involved and to provide a counterexample to S-W's assertion that public provision will never exceed the market equilibrium quantity.

I. Public Provision of Private Goods: The General Problem

As an alternative to market provision, public provision of a private good involves financing through some tax-sharing arrangement and distribution to citizens according to some distribution scheme. This may or may not be followed by private exchange of the publicly provided good.[3] If private exchange is allowed in the postdistribution period, then citizens in their roles as voters are faced with a complex choice problem in evaluating alternative proposals. The voter must form some expectation about outcomes at the exchange stage; in particular, about the value of his public allotment of the private good and its relative price at the exchange stage. Only with these expectations can he rank alternative levels of public provision against one another or against market provision.

These expectations, moreover, will be conditioned by the economic setting in the community. This follows since actual outcomes at the exchange stage depend critically, for example, on whether the community is an open and small part of a much larger economy or a closed economy. The voter's evaluation of a particular level of public provision will, therefore, depend upon the structure of the exchange market. An analysis of the likely outcome of the majority voting process must be based upon an explicit modeling of these expectations.

---

[2] In our notation, w_i would be greater than or equal to the individual's expenditure, at the exchange stage, on good z_i plus his expenditure on good q_i. The individual's income at the start of the next period in an endowment model of the type we constructed would be given by y_i not w_i. This assumes, of course, that there is no savings in the model. If one allows for saving, then w_i would constrain the sum of individual consumption expenditures plus savings. The individuals' income next period would then be given by y_i plus the returns from his savings. Individual utility functions would have to be modified appropriately to extend the model in this direction. The analysis of the majority-voting equilibrium also becomes more difficult in a multiperiod and, hence, multi-issue setting.

[3] If private exchange is prohibited then the analysis of Sections I and II of our original paper applies.

One natural approach to this problem, since uncertainty elements are ignored, is to assume that expectations are formed rationally, being based on perfect foresight. Analytically, the problem can then be solved by working backwards as follows. First, solve for the equilibrium outcomes of the exchange process resulting from alternative levels of public provision. In general, these outcomes will depend upon preferences, the distribution of income, tax shares, benefit shares, the structure of the exchange market, and cost conditions. Second, the equilibrium outcomes that result at the exchange stage from various levels of public provision can be substituted into the direct utility function for each voter to derive that voter's public sector preference function.[4] Third, these public sector preference functions, along with the indirect utility functions that result from private market provision, can be aggregated through majority voting.

Working with these public sector preference functions, one can examine the following questions. Does a majority rule equilibrium exist? Under what conditions will the equilibrium involve public provision of the private good? How likely are these conditions to be satisfied? And, if there is public provision, how will the level of provision compare to the level of provision in the private market? This approach is illustrated below and applied in the following section.

A. *The Private Market Equilibrium*

Consider a political community with m citizen-voters, each of whom has preferences representable by the following utility function:

$$(1) \qquad U_i = U_i(z, q_i) \qquad i = 1, \ldots, m$$

where z_i and q_i are the ith individual's consumption of two pure private goods.

Under private market provision of both z and q, the individual will maximize (1) subject to the following budget constraint:

$$(2) \qquad y_i \geqslant z_i + pq_i$$

where y_i is his initial income or endowment, z_i is taken as numeraire, and p is the price of q. This maximization problem yields individual demand functions of the form:

$$(3) \qquad \begin{aligned} q_i^d &= g_i(p, y_i) \\ z_i^d &= y_i - pg_i(p, y_i) \end{aligned}$$

Assuming that the unit cost of q, denoted c, is constant, then in market equilibrium, denoted by a superscript M:

$$(4) \qquad p^M = c \quad \text{or} \quad \sum_{i=1}^{m} g_i(c, y_i) = x^M$$

where x^M is the total amount of good q consumed in market equilibrium by the citizens of this community. The citizen-voter's indirect utility function evaluated at equilibrium prices and quantities provides him with an expression for ranking market provision against alternative levels of public provision and is given by

$$(5) \qquad U_i^M = U_i[\, y_i - cg_i(c, y_i), g_i(c, y_i)\,]$$

B. *Public Provision with Private Exchange*

1. *The Exchange Stage*

Assume that good q is now publicly provided while z remains privately provided. The publicly provided good is distributed among the voters according to the following scheme:

$$(6) \quad d_i = \alpha_i x$$

$$\text{where} \qquad \alpha_i \geqslant 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 1$$

---

[4] See James Barr and Otto Davis for a discussion of public sector preference functions.

where x is the total provision, α_i is voter i's benefit share or coefficient of distribution, and d_i is the amount of the good distributed to voter i. The ith voter's budget constraint at the exchange stage is given by

$$y_i + pd_i \geqslant z_i + \tau_i cx + pq_i \tag{7}$$

where τ_i is his tax share, c is the assumed constant marginal and average cost of a unit of x to the community, and p is the price of the publicly provided good at the exchange stage. Individual demands at the exchange stage are given by the maximization of (1) subject to (6), (7), and a given level of provision of x:

$$x = x^0 \tag{8}$$

This maximization problem yields individual demand functions of the form:

$$q_i^d = g_i(p, w_i) \tag{9}$$

$$z_i^d = w_i - pg_i(p, w_i)$$

$$\text{with} \qquad w_i \equiv y_i + (\alpha_i p - \tau_i c)x^0$$

In these functions, w_i is simply the individual's wealth at the exchange stage.

The equilibrium price determined at the exchange stage depends not only upon these individual demands and the level of public provision of x, but also upon the structure of the exchange market in the postdistribution setting. Several different structures for this market can be considered.

I. *The Exchange Market is Open*: The community is a small part of a larger economy with citizens able to purchase (import) additional units of q from individuals outside the community or sell (export) the units of q distributed to them to individuals outside the collectivity.

In this case, the equilibrium exchange price, denoted p^E, will be determined by supply and demand conditions in the rest of the economy and will thus be given by

$$p^E = c \tag{10}$$

II. *The Exchange Market is Closed*: The only q available at the exchange stage is that distributed by the government, so that there are neither imports, exports, nor any domestic production of q at the exchange stage.

In this case, the equilibrium exchange price depends solely on supply and demand conditions in the community and is determined by the following condition:

$$\sum_{i=1}^{m} g_i[p^E, y_i + (\alpha_i p^E - \tau_i c)x^0] = x^0$$

Solving this equation for the equilibrium exchange price gives

$$p^E = \psi(x; \boldsymbol{\alpha}, \boldsymbol{\tau}, \mathbf{y}, c) \tag{11}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$, and $\mathbf{y}$ are the vectors of benefit shares, tax shares, and incomes, respectively.

III. *The Exchange Market is Open to Imports But Closed to Exports*: At the exchange stage citizens can purchase additional units of q from outside the community or from domestic producers at a price of c. Resale of the publicly provided good to individuals outside the community is prohibited.

In this case,

$$p^E = \min[\psi(x), c] \tag{12}$$

IV. *The Exchange Market is Open to Exports But Closed to Imports*: At the exchange stage citizens can sell the units of the publicly provided good distributed to them to individuals outside the community at a price of c. Purchases from outsiders or domestic producers at the exchange stage is prohibited.

In this case,

$$p^E = \max[\psi(x), c] \tag{13}$$

The level of utility, denoted U_i^E, that the ith voter receives from a level of public provision of x^0, given a particular structure of the exchange market can be found by substituting the appropriate equation from equations (10)–(13) into the exchange-stage demand functions, equations (9), to determine the equilibrium levels of consumption, denoted z_i^E and q_i^E, and then substituting these consumption levels into the direct utility function. That is,

$$(14) \qquad U_i^E = U_i(z_i^E, q_i^E)$$

$$\text{where} \qquad q_i^E = g_i(p^E, w^E)$$

$$z_i^E = w_i^E - p^E g_i(p^E, w_i^E)$$

$$w_i^E = y_i + (\alpha_i p^E - \tau_i c)x^0$$

and p^E is given by the appropriate expression from equations (10)–(13) depending upon the structure of the exchange market.[5]

2. *The Voting Stage*

A voter's public sector preference function provides a ranking for that voter of alternative states of the public sector. Under the assumption of perfect foresight, equation (14) provides the appropriate ranking. At the voting stage, the voter's public sector preference function, denoted by $V_i(x)$, can be used for evaluating alternative proposals for levels of public provision. Moreover, by comparing equation (5) to (14), the voter can also evaluate market provision relative to various levels of public provision.

So long as private market provision is an allowable proposal, the following two conditions must be satisfied if a proposed level of public provision of the private good x^* is to be an equilibrium of the majority voting process:

(i) x^* must be an element of the set X defined as the set of all x such that $s(x) - l(x) \geqslant 0$ where $s(x)$ is the cardinality of the set of voters who strictly prefer that level of public provision to market provision (i.e., $V_i(x) > U_i^M$) and $l(x)$ is the cardinality of the set of voters who strictly prefer market provision to that level of public provision (i.e., $V_i(x) < U_i^M$).

(ii) for any $x \epsilon X$, $n(x^*, x) \geqslant n(x, x^*)$ is the cardinality of the set of voters for whom $V_i(x^*) > V_i(x)$ and $n(x, x^*)$ is the cardinality of the set of voters for whom $V_i(x) > V_i(x^*)$.

If the set X is empty, then private market provision will be the equilibrium of the majority voting process.

By considering specific forms of the direct utility function, and thus specific forms of equation (14), we can examine conditions under which an equilibrium with public provision will exist and the characteristics of public provision relative to private provision.

II. A Counterexample

The framework presented above provides an excellent means of evaluating S-W's contribution. Consider the following summary of their analysis: "...the principle remains that the public provision of a private good will be identical to market provision after incorporating the income effects of different tax prices and the postredistributional effect of exchange" (p.464). This is either a remarkable result or a tautology, depending upon what is meant by their qualifying phrase. To show that this assertion is incorrect, we have chosen as our counterexample a special case in which the income elasticity of demand for the publicly provided good is zero, so that income effects are absent. This counterexample also demonstrates that public provision may well exceed the level of private market provision. We consider both the open-economy case, in order to correct an error in our earlier paper, and the closed-economy case.

A. *The Private Market Equilibrium*

Assume that all the citizen-voters have identical preferences given by

$$(15) \qquad U_i = z_i + \left(\gamma q_i - \frac{\beta}{2} q_i^2\right), \; \gamma, \beta > 0$$

This utility function yields a linear demand function for q that is independent of the level of income. That is,

$$(16) \qquad q_i^d = \frac{1}{\beta}(\gamma - p)$$

---

[5] It is important to note that in general voter i's evaluation of a level of public provision depends upon the distribution of benefit shares, tax shares, and income since these distributions may effect the equilibrium exchange price.

In market equilibrium, the following conditions hold:

$$(17)\qquad p^M = c \text{ or } x^M = \frac{m}{\beta}(\gamma - c)$$

In private market equilibrium, the ith citizen receives a level of utility given by

$$(18)\qquad U_i^M = y_i + \frac{1}{2}\,\frac{(\gamma - c)^2}{\beta}$$

B. *Public Provision with Private Exchange*

1. *The Exchange Stage*

At the exchange stage, individual demands are given by

$$(19)\qquad q_i^d = \frac{1}{\beta}(\gamma - p)$$

$$z_i^d = y_i + (\alpha_i p - \tau_i c)x^0 - \frac{p}{\beta}(\gamma - p)$$

Given these individual demands, the equilibrium exchange price is given by

$$(20)\qquad p^E =$$

Open Economy

$$c$$

Closed Economy

$$\begin{array}{ll} \gamma - \dfrac{\beta}{m}x^0 & \text{for } x^0 \leqslant \dfrac{m\gamma}{\beta} \\ 0 & \text{for } x^0 > \dfrac{m\gamma}{\beta} \end{array}$$

Open To Imports/ Closed To Exports

$$\begin{array}{ll} c & \text{for } x^0 < x^M \\ \gamma - \dfrac{\beta}{m}x^0 & \text{for } x^M \leqslant x^0 < \dfrac{m\gamma}{\beta} \\ 0 & \text{for } x^0 > \dfrac{m\gamma}{\beta} \end{array}$$

Open To Exports/ Closed To Imports

$$\begin{array}{ll} \gamma - \dfrac{\beta}{m}x^0 & \text{for } x^0 < x^M \\ c & \text{for } x^0 \geqslant x^M \end{array}$$

Note that in this special case, the ψ function is independent of $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$, and $\mathbf{y}$.

2. *The Voting Stage*

Open Economy. If the exchange market is open, then p^E equals c and substituting this into equations (19), (19) into (15), and rearranging gives the following public sector preference function:

$$(21)\qquad V_i = U_i^M + (\alpha_i - \tau_i)cx \equiv V_i(x; \alpha_i\tau_i, y_i, c)$$

$$\text{with}\quad \frac{\partial V_i}{\partial x} = (\alpha_i - \tau_i)c \quad\text{and}\quad \frac{\partial^2 V_i}{\partial x^2} = 0$$

The voters' public sector preference functions are linear with an intercept of U_i^M and a slope of $(\alpha_i - \tau_i)c$. Note that whether the voter's utility increases or decreases with increases in x depends solely on his benefit share relative to his tax share. That is,

$$\frac{\partial V_i}{\partial x} \overset{<}{\underset{>}{=}} 0 \quad\text{as}\quad \alpha_i \overset{<}{\underset{>}{=}} \tau_i$$

In this case, then, if a voter's tax share is less than his benefit share, he will vote for unbounded increases in the level of x; if his tax share is greater than his benefit share, his most preferred level of x is zero; and if his tax share equals his benefit share, he is indifferent between market provision and *any* level of public provision.

With this formulation of the problem, it is quite easy to examine the conditions under which public provision at some positive level will be an equilibrium of the majority voting process. If a majority of voters have benefit shares strictly greater than their tax shares, then this coalition will support unbounded increases in the level of public provision. Public provision under these conditions would imply bankruptcy for the minority coalition, and hence is unlikely to continue. This extreme result can be avoided in a number of ways. The first way is to restrict the distribution of tax and benefit shares so that this condition cannot arise. For example, if there are three voters, A, B, and C, and $\tau_A > \alpha_A, \tau_B = \alpha_B$, and $\tau_C < \alpha_C$, respectively, *any* level of public provision will be an equilibrium of the voting process. This example also clearly indicates that it is not necessary for a *majority* of voters to have benefit shares *identically* equal to their tax shares for finite levels of public provision to be equilibria. Unfortunately, we mistakenly stated the condition in this strong form in our earlier paper, and thus welcome this opportunity to correct our error.

A second way to eliminate the extreme result of an infinite level of provision is to restrict the proposals that are allowed on the agenda, as suggested by S-W. The restriction they place on allowable proposals—that proposed levels of spending must be less than or equal to the minimum value of $y_i/\tau_i c$ across *all* voters—is, however, not particularly compelling, as we discussed earlier. Other restrictions can be easily imagined, but appeal to these types of restrictions, in our opinion, only strengthens the argument that public provision of easily resalable private goods in an open-economy setting is unlikely.

Closed Economy. By restricting the model somewhat more, we are able to show that even in the absence of income effects there is no necessity for the equilibrium level of public provision to coincide with, or be less than, the level of market provision. Assume that benefit shares are equal across all voters; i.e., $\alpha_i = 1/m$. The voter's public sector preference function with equal sharing and a closed-exchange market can be found by substituting this information along with (20) into (19), then (19) into (15), and rearranging to yield:

(22)

$$V_i(x) = \begin{cases} y_i + \dfrac{1}{m}(\gamma - m\tau_i c)x - \dfrac{1}{2}\dfrac{\beta}{m^2}x^2 & \text{for } x \leqslant \dfrac{m\gamma}{\beta} \\ y_i + \dfrac{1}{2}\dfrac{\gamma^2}{\beta} - \tau_i c x & \text{for } x > \dfrac{m\gamma}{\beta} \end{cases}$$

This function is single peaked for $0 < \tau_i < 1$. For values of τ_i greater than or equal to γ/mc, the peak occurs at zero. For values of τ_i less than γ/mc and greater than zero, the peak occurs at a level of provision denoted $\hat{x}_i$, given by

$$\hat{x}_i = \frac{m}{\beta}(\gamma - m\tau_i c) \tag{23}$$

For τ_i equal to zero, the voter's public sector preference function reaches a peak at $m\gamma/\beta$ and is flat for further increases in x. Other important properties of these public sector preference functions are summarized in Table 1.

TABLE 1

τ_i	$\hat{x}_i$	$V_i(\hat{x}_i)$ Relative to U_i^M
$\tau_i > \frac{1}{m}\frac{\gamma}{c}$	$\hat{x}_i = 0$	$V_i(\hat{x}_i) < U_i^M$
$\frac{1}{m} < \tau_i < \frac{1}{m}\frac{\gamma}{c}$	$0 < \hat{x}_i < x^M$	$V_i(\hat{x}_i) < U_i^M$
$\tau_i = \frac{1}{m}$	$\hat{x}_i = x^M$	$V_i(\hat{x}_i) = U_i^M$
$\tau_i < \frac{1}{m}$	$\hat{x}_i > x^M$	$V_i(\hat{x}_i) > U_i^M$
$\tau_i = 0$	$\hat{x}_i = \frac{m\gamma}{\beta}$	$V_i(\hat{x}_i) > U_i^M$

This example makes it quite clear that S-W's assertions about public provision relative to private provision are incorrect. Consider the case they examined, a proportional income tax and a nonsymmetric income distribution, for which the median income is less than the mean income. In this case, a majority of voters have tax shares strictly less than $1/m$. It can be shown that the equilibrium of the majority voting process is given by the median distribution of $\hat{x}_i$s, denoted $\hat{x}_{MED}$, and as Table 1 indicates, this level of provision *exceeds* the level of private market provision.[6] This example, in fact, yields a conclusion just the *opposite* of S-W's: if there is public provision, it will be at a level greater than or equal to the level of private provision.[7]

---

[6] That is, it can be shown that for a majority of voters $V_i(\hat{x}_{MED}) > U_i^M$ and $n(\hat{x}_{MED}, x) > n(x, \hat{x}_{MED})$ for all other $x \in X$.

[7] Other market structures are also possible. If the exchange market is open to exports but closed to imports, and a majority of voters have tax shares less than $1/m$, the equilibrium is the same as that described for the open-economy case—unbounded increases in the proposed level of provision. If the exchange market is

Other cases are of course possible. For example, if tax shares instead of benefit shares are equal, if demand in the market equilibrium is inelastic (in the sense that γ is greater than $2c$), and if a majority of voters have benefit shares greater than $1/m$, then the majority-voting equilibrium will exhibit positive public provision, but with less than the market level of provision. The general problem laid out in the previous section, therefore, does not lead to the strong results suggested by S-W. Appeals to the Coase Theorem provide little support for such strong assertions, since the political environment being examined is one of majority voting with given fiscal institutions. There is no direct bargaining or vote trading.

---

open to imports but closed to exports, and a majority of voters have tax shares less than $1/m$, then the equilibrium is the same as that described for the closed-economy case. These two cases moreover lead to an important observation. If the exchange market is closed to exports solely because of legal prohibitions imposed by the collectivity, then the equilibrium described above for the closed-economy case is subject to being overturned by a proposal to eliminate these prohibitions on resale.

REFERENCES

J. Barr and O. Davis, "An Elementary Political and Economic Theory of the Expenditures of Local Governments," *Southern Econ. J.*, Oct. 1966, *33*, 149–65.

A. T. Denzau and R. J. Mackay, "Benefit Shares and Majority Voting," *Amer. Econ.Rev.*, Mar. 1976, *66*, 69–76.

R. J. Staaf and E. G. West, "Limits on the Public Provision of Private Goods," *Amer.Econ. Rev.*, June 1980, *70*, 461–66.

The Interpretation of Dummy Variables in Semilogarithmic Equations

By Robert Halvorsen and Raymond Palmquist*

A number of recent articles on discrimination, education, and income misinterpret the regression of coefficients of some variables, and as a result report incorrect estimates of their effects. Examples appearing in this *Review* include articles by Eric Hanushek and John Quigley, David Hartman, Robert Lucas, James Smith and Finis Welch, Barry Chiswick, and Emily Hoffman. An analogous error appears in the literature on hedonic price indexes, including frequently cited articles by Zvi Griliches, Jack Triplett, and A. T. Court.

This common error involves the interpretation of the coefficients of dummy variables in semilogarithmic regression equations. The articles cited above assume that the coefficient of a dummy variable, multiplied by 100, is equal to the percentage effect of that variable on the variable being explained. However, it is easily shown that this interpretation, while correct for continuous variables, is not correct for dummy variables and can result in substantial errors in the reporting of results.

The general form of the equations estimated in the articles cited above is

$$(1) \qquad ln\, Y = a + \sum_i b_i X_i + \sum_j c_j D_j$$

where the X_i represent continuous variables and the D_j represent dummy variables. The coefficient of a continuous variable is

$$b_i = \frac{\partial ln\, Y}{\partial X_i} = \frac{1}{Y} \cdot \frac{\partial Y}{\partial X_i}$$

Thus the coefficient of a continuous variable, multiplied by 100, is equal to the percentage effect on Y of a small change in that variable.

Since a dummy variable enters the equation in dichotomous form, the derivative of the dependent variable with respect to the dummy variable does not exist. Instead, the coefficient of a dummy variable measures the discontinuous effect on Y of the presence of the factor represented by the dummy variable. The appropriate interpretation of the coefficient of a dummy variable can be shown directly by a transformation of equation (1).

For simplicity, it will be assumed that there is a single dummy variable. Equation (1) can then be written as

$$(2) \qquad Y = (1+g)^D \exp\left(a + \sum_i b_i X_i\right)$$

where g is the relative effect on Y of the presence of the factor represented by the dummy variable.[1] Thus the coefficient of the dummy variable in equation (1) is $c = ln(1+g)$. The relative effect on Y is $g = \exp(c) - 1$, and the percentage effect is equal to[2]

$$100 \cdot g = 100 \cdot \{\exp(c) - 1\}$$

The results reported in the studies cited above are based on the incorrect assumption that $c = g$. The relationship between c and g when the absolute value of g is less than one can be examined by expansion of c,

---

*University of Washington. We are grateful to Taylor Dennen and Gerald Glandon for discussion of the topic of this note.

[1] Thus $g = (Y_1 - Y_0)/Y_0$ where Y_1 and Y_0 are the values of the dependent variable when the dummy variable is equal to one and zero, respectively.

[2] When time dummies are used to estimate hedonic price indexes, c is equal to the rate of change in price which, when continuously compounded, yields g, the relative change in price during the period. The value of the price index at the end of the period is equal to exp (c), not $1+c$ as assumed in the hedonic studies cited above.

$$c = ln(1+g) = g - \frac{1}{2}g^2 + \frac{1}{3}g^3 - \dots \quad |g| < 1$$

For small values of g, c is approximately equal to g. When g is positive, c is smaller than g, and when g is negative, c is algebraically smaller than g but larger in absolute value.

Table 1 shows the magnitude of the relative change g, implied by selected values for the coefficient of the dummy variable c. The errors involved in assuming that c is equal to g can be substantial for values of c within the range estimated in the cited studies. For example, Hanushek and Quigley (p. 74) report that a postgraduate degree increases the wages of a black worker by 64 percent, whereas the correct result implied by their regression is 90 percent.

TABLE 1—RELATIONSHIP BETWEEN DUMMY VARIABLE COEFFICIENTS AND RELATIVE EFFECTS

Coefficient of the Dummy Variable (c)	Relative Effect (g)
1.50	3.48
1.25	2.49
1.00	1.72
0.75	1.12
0.50	0.65
0.25	0.28
0.00	0.00
−0.25	−0.22
−0.50	−0.39
−0.75	−0.53
−1.00	−0.63
−1.25	−0.71
−1.50	−0.78

REFERENCES

B. R. Chiswick, "Sons of Immigrants: Are They at an Earnings Disadvantage?," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 376–80.

A. T. Court, "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, New York 1939.

Z. Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in his *Price Indexes and Quality Change*, Cambridge 1971.

E. A. Hanushek and J. M. Quigley, "Implicit Investment Profiles and Intertemporal Adjustments of Relative Wages," *Amer. Econ. Rev.*, Mar. 1978, *68*, 67–79.

D. G. Hartman, "What Do Economics Majors Learn?," *Amer. Econ. Rev. Proc.*, May 1978, *68*, 17–22.

E. P. Hoffman, "Faculty Salaries: Is There Discrimination by Sex, Race, and Discipline? Additional Evidence," *Amer. Econ. Rev.*, Mar. 1976, *66*, 196–98.

R. E. B. Lucas, "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *Amer. Econ. Rev.*, Sept. 1977, *67*, 549–58.

J. P. Smith and F. R. Welch, "Black-White Male Wage Ratios: 1960–70," *Amer. Econ. Rev.*, June 1977, *67*, 323–38.

J. E. Triplett, "Automobiles and Hedonic Quality Measurement," *J. Polit. Econ.*, May/June 1969, *77*, 408–17.

Production, Information Costs, and Economic Organization: The Buyer Monitoring Case

By NANCY L. JACOB AND ALFRED N. PAGE*

In their December 1972 article in this *Review*, Armen Alchian and Harold Demsetz (hereafter, A-D) provide the conditions which generate the classical capitalistic firm.[1] In such a firm, team production is monitored by a residual claimant to the firm's earnings. The monitor measures input productivity (the degree of shirking) and meters rewards for that productivity. The need for monitoring arises, as in their case of two men jointly lifting heavy cargo into trucks, because individual marginal productivity cannot be determined by observing team output.

The explanatory power of the A-D thesis can be increased by noting that the firm is composed of a number of such teams, each having a readily identifiable physical output as well as a relation to demand that is more troublesome to measure. The job then for the residual claimant is to determine the size and makeup of his various teams and to monitor his team of teams. Since the monitor cannot simultaneously watch each team's efforts, and since the nature of each team's work will differ, the residual claimant will develop a variety of schemes to aid in the monitoring process. This note analyzes the economics of a scheme of particular interest, the utilization of the buyers of the firm's production as monitors.

I

The interaction between an employee and a buyer can affect the buyer's demand for the production of the rest of the firm's teams. For example, an unpleasant or uninformed salesman can lower the desirability of a durable good and a surly waiter can lower the desirability of a meal. Furthermore, just as an A-D employee will shirk in a factory to maximize private gains if left unmonitored, employees and buyers will collude for their own private benefits if left unmonitored. Buyers will offer employees remuneration in exchange for goods priced at less than their full opportunity costs, and employees will agree as long as they do not bear those opportunity costs.

Since every interaction between an employee and a buyer is unique, monitoring these interactions presents the residual claimant with a difficult problem. The residual claimant cannot watch over every transaction personally, and accounting and statistical tools, being dependent on regularity in data, become of limited value in assessing this part of an employee's performance. Thus while owners may be able to institute systems to minimize employee fraud and absenteeism, owners cannot efficiently observe the knowledge, politeness, and deference which employees exhibit when interacting with buyers. As a result, an optimizing residual claimant may find that it is economic to monitor only a portion of such an employee's work. For the remaining portion, residual claimants will develop techniques to take advantage of the buyer's interest in monitoring employee behavior.

II

We can gain insight into the use of buyer monitoring by examining a model of the firm. We begin the analysis by making the following simplifying assumptions:

1. The firm produces a single, differentiated product having two input factors: factor 1 (a portion of labor input engaged in the team production of revenue which interacts with buyers); and factor 2 (all other inputs). Institutionally, factor 1 can be

---

*University of Washington.

[1] For an analogue to the A-D approach in the agency literature, the reader should see Michael Jensen and William Meckling.

monitored either by the firm's owners (for example, by a residual claimant) or by the buyers of the firm's production, while factor 2 is monitored solely by the residual claimant. The markets for factor 1 and factor 2 are assumed perfect.

2. Buyers and owners have constant, though possibly unequal, costs of monitoring per unit of factor 1 employed by the firm. All buyers are viewed, for simplicity, as identical to one another.

3. The productivity of, as well as the remuneration received by, the suppliers of factor 1 may depend on the degree of monitoring activity engaged in by owners or buyers and on the system of monitoring chosen.

Using these assumptions, the total output (Q) of the firm will be governed by the quantities of factor 1(X_1) and of factor 2(X_2), as well as on the level of monitoring activity (δ_0) chosen by the firm. If the owners thus monitor factor 1 and if γ represents their joint, constant-monitoring cost per unit of factor 1, then the firm's profit function can be written as

$$\text{(1)} \qquad \pi_0 = P\cdot Q(X_1, X_2, \delta_0) - (W_1(\delta_0)+\gamma)X_1 - W_2X_2$$

where P = the price per unit of output and W_1, W_2 = the remuneration paid per unit of X_1 and X_2, respectively. Note that in (1) the suppliers of factor 1 receive a remuneration W_1 which depends on the level of monitoring activity the owners engage in.

If buyers are utilized by owners to monitor factor 1, and if buyers have constant monitoring costs (θ) per unit of factor 1, the firm's profit function can be written as

$$\text{(2)} \qquad \pi_B = \left(P - \frac{(\eta+\theta)X_1}{Q}\right)Q(X_1, X_2, \delta_B) - W_2X_2$$

where δ_B is the level of monitoring activity chosen by the buyers.

This function reflects the fact that the suppliers of factor 1 now are receiving remuneration (η) that is associated with buyer monitoring. Depending on institutional considerations, η may be paid either in the form of "tips" or "commissions."[2] In turn, the price per unit of output received by the firm, which is available for profits and returns to other inputs, is reduced by the amount of this remuneration as well as by buyers' implicit monitoring costs.

If π_0 exceeds π_B, the owners will maximize profits by delegating monitoring activities associated with factor 1 to a residual claimant; if the reverse is true, buyers will be preferred as monitors. In general, the relationship between π_0 and π_B, and, therefore, the firm's optimal choice of monitors, depends on:

(a) The extent to which buyers monitor more or less effectively than do owners (i.e., the extent to which buyers choose a δ_B that differs from δ_0). Since the firm's output, holding X_1 and X_2 constant, is a function of δ, the effect of differences in monitoring effectiveness among owners and buyers may be to alter the optimal factor mix as well as the optimal price and output decisions of the firm.

(b) The difference in monitoring costs between owners and buyers (i.e., the relative magnitudes of γ and θ). The impact of (b) can be most clearly seen by supposing that buyers and owners choose the same level of monitoring activity and that total payments to factor 1 per unit of that factor are the same no matter who monitors ($\eta = W_1$). In that case, assuming that the firm's optimal price and output decisions are the same under both monitoring systems, as one might expect, the necessary and sufficient condition for the optimality of buyer monitoring is $\theta < \gamma$. Further, if the firm's optimal price and output decisions *do* depend upon the choice of the monitoring system, it must a fortiori be true that if $\theta < \gamma$, buyer monitoring will be preferred to owner monitoring of factor 1.

---

[2] Our search of the traditional economics literature turned up only four references to tipping, none of which referenced the others or provided an analytical model of the practice. S. N. S. Cheung in a footnote observed that "'tipping' payments (as in the case of waitresses) are preferred to other forms [of policing employees] when the quality of services are significant" (p. 68). Kenneth Brown invited readers to provide an explanation of tipping in the "Puzzles and Problems" section of the *Journal of Political Economy*. This section was discontinued before an answer appeared. John Pencavel recognized that tipping implied that work performance is evaluated directly by consumers. David Hemenway discusses many of the current institutional details of tipping.

Finally, of course, the optimal choice of monitoring system depends simultaneously on both (a) and (b) and thus depends on the elasticity of the firm's demand curve as well as upon its production technology.

III

An implication of the model is that certain firms will utilize mixed-monitoring systems, that is, systems where both buyers and owners are engaged in the supervision of employees. Institutional examples are where supervisors hired by owners monitor commission sales clerks, and where waiters are monitored and paid by owners as well as monitored and tipped by buyers.[3]

Two other testable implications of the model are of interest. First, a buyer of the team production should not find it in his interest to tip an owner of a firm, the residual claimant. This implication is completely consistent with institutional tipping conventions. Second, if our extension of the model is correct, the practice of tipping and the utilization of commissions should have occurred as long as the A-D firm has existed. While the history of tipping is far from complete and there is not unanimity even with respect to the origin of the work, there is no question that the practice, as defined in this paper, has existed as long as the capitalistic firm.[4] With respect to commissions, their use as a means of payment can be traced back to mercantile days.

In summary, buyer monitoring has been and always will be an integral part of the Alchian-Demsetz firm.

---

[3] For more specific examples, see U.S. Bureau of Labor Statistics.

[4] As just one example from early English literature, George Farquhar in his 1707 play, *Beaux Strategem*, wrote, "Then I, Sir, tips me the verger with half-a crown."

REFERENCES

A. A. Alchian and H. Demsetz, "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*, 777–95.

K. Brown, "Puzzles and Problems, *Q* 722," *J. Polit. Econ.*, Jan./Feb. 1972, *80*, 213.

S. N. S. Cheung, *The Theory of Share Tenancy*, Chicago 1969.

David Hemenway, *Prices and Choices: Microeconomic Vignettes*, Cambridge 1977.

M. Jensen and W. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure," *J. Finan. Econ.*, Oct. 1976, *3*, 305–60.

J. H. Pencavel, "Work Effort, On-the-Job Screening, and Alternative Methods of Remuneration," in Ronald G. Ehrenberg, ed., *Research in Labor Economics*, Vol. 1, Greenwich 1977, 231.

U. S. Bureau of Labor Statistics, "Wages and Tips in Restaurants and Hotels, March 1970," Bull. 1712, Washington 1971.

Paglin's Gini Measure of Inequality: A Modification

By JOHN P. FORMBY AND TERRY G. SEAKS*

In a recent issue of this *Review*, Morton Paglin made an important contribution to the literature on the measurement and trend in income inequality. It has long been recognized that the Lorenz curve and associated Gini coefficient (L-Gini) ignores life cycle effects on income, and consequently is deficient as a measure of income inequality.[1] Paglin's contribution lies in the development of a basic revision of the Lorenz curve from which a new Gini measure of inequality is constructed. His new line of income equality (the *P* curve) excludes interage inequality. Paglin intended the *P* curve to reflect equal lifetime incomes adjusted for the life cycle effects inherent in the age-income profile (1975, pp. 599–601). Paglin uses the deviations of the actual income distribution away from the *P* curve to construct the Paglin-Gini (P-Gini) which is used as his measure of income inequality. Paglin's contribution is important because he is able to adjust the income distribution for the age-income profile while retaining a single summary measure of income inequality.

A number of comments, criticisms, and suggested extensions of Paglin's work have appeared in this *Review*.[2] Perhaps the most telling criticism is that the *P* curve does not represent a precise adjustment of the income distribution for the age-income profile and the age composition of the population.[3] Since the age-cohort income distributions overlap, the *P* curve can only be interpreted as reflecting an imperfect adjustment of the income distribution for life cycle effects. However, when compared to the Lorenzian line of equality, which ignores all life cycle effects, the *P* curve is a superior base from which to measure deviations of the actual income distribution from equality.

In this note we accept Paglin's contention that the P-Gini is an improvement over the L-Gini, but argue that he fails to take his insight to its logical conclusion. We present a modified Paglin-Gini (MP-Gini) which in terms of Paglin's own logic is more defensible than the P-Gini. Our MP-Gini, like the P-Gini, shows substantial declines in income inequality in the United States since 1947. However, the MP-Gini suggests a more modest decline than does the P-Gini. We also show that the MP-Gini, like the L-Gini and unlike the P-Gini, has the conceptually desirable property of always varying on the 0–1 interval. Nelson in his criticism of Paglin notes that a problem exists with the interval on which the P-Gini varies (p. 498), but fails to correctly specify the interval over which it does vary.

The distinction between the L-Gini, the P-Gini, and the MP-Gini is made clear by reference to Figure 1 which is equivalent to Paglin's Figure 1B. The *E* curve is the line of absolute equality and the *L* curve is the standard Lorenz curve. Assuming family income is equally distributed within an age class, Paglin's *P* curve is constructed for a given year by ranking age groups in ascending order by mean incomes and cumulating in a fashion analogous to the Lorenz curve. Given the parabolic age-income profile and the assumption that incomes are equally distributed within an age cohort, the *P* curve shows a perfectly equal income distribution. Deviations of the Lorenz curve away from the *P* curve reflect income inequality that cannot be traced to the underlying age-income profile. The area α shows that part of the Lorenzian inequality which is traceable to the age-income profile. The area β shows the Lorenzian inequality which is not attributable to the age-income profile. The area γ can be thought of as the area of equality. The standard L-Gini is $(\alpha+\beta)/(\alpha+\beta+\gamma)$.[4] Paglin defines the P-Gini as follows: "The Lorenz-Gini minus the age-Gini equals the Paglin-Gini" (pp. 600–01). This is equivalent to defining the P-Gini as $\beta/(\alpha+\beta+\gamma)$. The problem with this procedure is that the area α reflects the effect of the age-income profile and presumably has only mathematical significance and is thus not relevant in measuring income inequality. Paglin demonstrates this. Indeed, it is a primal point of his paper and the reason he discards α in the numerator of the P-Gini. However, he has left α in the denominator. If α is deleted from the numerator, why should it be retained in the denominator? There appears to be no strong reason for keeping it and some good reasons for discarding it.

---

*Professor and associate professor, respectively, University of North Carolina-Greensboro.

[1] James Morgan is emphatic on this point, pp. 273–75.

[2] See the comments by Eric Nelson; William Johnson; Sheldon Danziger, Robert Haveman, and Eugene Smolensky; Joseph Minarik; and C. John Kurien as well as Paglin's reply.

[3] Nelson's and Johnson's comments both relate to this point. Nelson decomposes the Paglin-Gini and shows that it does not represent a precise correction of life cycle effects. Johnson's comment shows that the P-Gini does not precisely measure the extent of permanent lifetime income inequality and that Paglin's Gini always underestimates permanent income inequality.

We define the modified Paglin-Gini by removing α in both the numerator and denominator so that MP-Gini $=\beta/(\beta+\gamma)$. Figure 1 shows the values of the three Gini measures for 1972. The data in Paglin's (1975) Table 3 was used to compute MP-Ginis for the period 1947–72. Our Table 1 shows the annual values of the three Gini measures as well as the values of α, β, and γ from which they are computed. It is clear that the three Gini measures would be equal if the value of α were zero. A positive value of α leads to the relation:

$$\text{L-Gini} > \text{MP-Gini} > \text{P-Gini} \tag{1}$$

By eliminating α completely in the measurement of inequality, the MP-Gini measures the ratio of the deviation of the actual income distribution from the P reference curve *relative to the income which could be equally distributed given the age-income profile.* The P-Gini, on the other hand, measures the ratio of the deviation of the actual income distribution away from the P reference curve *relative to the income which could be equally distributed if the age-income profile were perfectly flat.* By including α in the denominator, the P-Gini understates income inequality and overstates the decline in inequality through time. This is shown clearly in Table 1; the P-Gini dropped from .303 in 1947 to .239 in 1972—a decline of 21 percent. The MP-Gini decreased from .328 to .272, indicating a 17 percent drop in inequality. To measure the compound rate of change in inequality over the period 1947–72, a least squares semilogarithmic growth curve was fitted to the P- and MP-Ginis. Inequality as measured by the P-Gini was found to be decaying at an annual compound rate of 1.0 percent. The MP-Gini was decaying at an annual rate of 0.8 percent.[5]

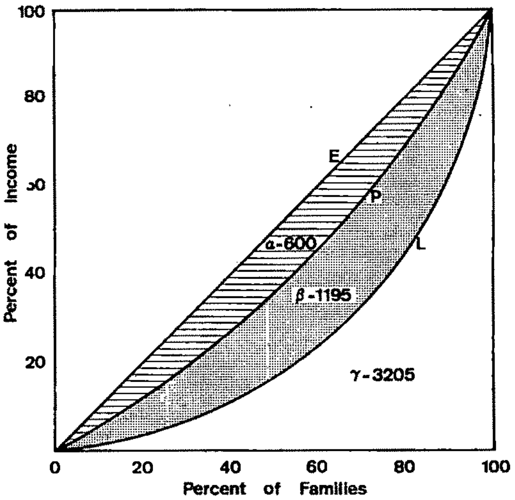


FIGURE 1. 1972 INCOME DISTRIBUTION

$$\text{L-Gini}=\frac{\alpha+\beta}{\alpha+\beta+\gamma}=.359 \qquad \text{A-Gini}=\frac{\alpha}{\alpha+\beta+\gamma}=.120$$

$$\text{P-Gini}=\frac{\beta}{\alpha+\beta+\gamma}=.239 \qquad \text{MP-Gini}=\frac{\beta}{\beta+\gamma}=.272$$

---

[4] In the standard Lorenz curve analysis there is of course no distinction between α and β. The area $\alpha+\beta$ is simply denoted as A, γ is labeled B, and the L-Gini is $A/(A+B)$.

[5] The compounded annual decay rate of the P-Gini was 1.044 percent and that of the MP-Gini was 0.839

TABLE 1—THREE DIFFERENT GINIS AND THEIR COMPONENTS: 1947–72

Year (1)	L-Gini (2)	P-Gini (3)	MP-Gini (4)	α (5)	β (6)	γ (7)
1947	.378	.303	.328	375	1515	3110
1948	.369	.293	.317	380	1465	3155
1949	.379	.303	.328	380	1515	3105
1950	.375	.298	.323	385	1490	3125
1951	.361	.289	.311	360	1445	3195
1952	.374	.296	.321	390	1480	3130
1953	.360	–	–	–	–	–
1954	.373	.289	.316	420	1445	3135
1955	.366	.278	.305	440	1390	3170
1956	.355	.272	.297	415	1360	3225
1957	.351	.262	.288	445	1310	3245
1958	.354	.261	.288	465	1305	3230
1959	.366	.276	.303	450	1380	3170
1960	.369	.278	.306	455	1390	3155
1961	.376	.286	.314	450	1430	3120
1962	.365	.265	.294	500	1325	3175
1963	.360	.262	.291	490	1310	3200
1964	.352	.256	.283	480	1280	3240
1965	.357	–	–	–	–	–
1966	.348	–	–	–	–	–
1967	.355	.245	.275	550	1225	3225
1968	.343	.233	.262	550	1165	3285
1969	.348	.233	.263	575	1165	3260
1970	.355	.241	.272	570	1205	3225
1971	.356	.237	.269	595	1185	3220
1972	.359	.239	.272	600	1195	3205

Source: Columns (2) and (3) come from Paglin (1975, Table 3). The remaining columns were computed based on information contained in Paglin's Table 3.

If the age-income profile never changed (α constant), the P-Gini and MP-Gini would decay at the same rate. As is obvious from Table 1, α is highly variable and has a pronounced upward time trend. Indeed, α has far more variability than either β or γ. Expressing this variability as a coefficient of variation, the respective coefficients for α, β, γ are .163, .081, and .016. Thus, α has twice the variability of β and ten times the variability of γ. The variability of α has important implications for the behavior of any type of Gini coefficient. The L-Gini is not affected by changes in α at all while the P-Gini is greatly affected. The MP-Gini is affected to a lesser degree. Over the period 1947–72, the mean elasticities with respect to α of L-Gini, MP-Gini and P-Gini are 0, –.243, and –.346, respectively.

A particularly interesting and noteworthy difference in the three Gini measures is the interval upon which they are bounded. As is well known, the L-Gini is bounded on the interval 0–1. This is true irrespective of the cross-sectional age-income profile prevailing at a moment in time and irrespective of changes in the profile through time. The property of being bounded on the 0–1 interval has certainly contributed to the intuitive appeal and widespread use of the L-Gini. The MP-Gini is also bounded on the 0–1 interval. This too is true irrespective of the age-income profile. The P-Gini, on the other hand, is bounded on the interval 0 and $\beta/(\alpha+\beta)$. If β goes to zero, then the P-Gini is zero. On the other hand, if β expands and γ goes to zero, then the P-Gini would take

---

percent. To test for a statistically significant difference in these decay rates, a test on the slopes of the semilogarithmic growth curves was performed. The hypothesis that the slopes were equal could be rejected at the .10 level, but the hypothesis was not rejected at the .05 level.

on its maximum value of $\beta/(\alpha+\beta)$. The P-Gini could reach a value of 1.0 only if $\alpha=0$. Recall that if α is zero, the P curve is the 45 degree line and the P-Gini, MP-Gini, and L-Gini are all equivalent. In practice, as Paglin has so ably pointed out, α is never zero, hence the P-Gini is not bounded on the 0–1 interval.

Nelson, in his criticism of Paglin, noted a problem with the interval over which the P-Gini varies. According to him, "The Paglin measure is bounded above by the [L-]Gini rather than unity" (p.498). It is true that if $\alpha=0$, the P-Gini equals the L-Gini but, as is well known, the L-Gini has an upper boundary of unity. The proper way to specify the upper boundary of the P-Gini is to let β expand and γ contract to the point that $\gamma=0$. Clearly the upper boundary is $\beta/(\alpha+\beta)$.

The P-Gini compared to the MP-Gini does have the characteristic of a constant denominator $(\alpha+\beta+\gamma)$ and thus has the convenient relation of differing from the L-Gini only by the age-Gini. This facilitates easy mental comparisons of the P- and L-Ginis. However, in our mind, the most important thrust of Paglin's paper is that the P curve is the relevant reference curve for measuring inequality and that the 45 degree line and the area α are not relevant. The use of α in the denominator of the P-Gini results in a measure which is not bounded on the 0–1 interval and which overstates the decline of income inequality. Even with these shortcomings and taking account of earlier criticisms, the P-Gini remains a useful measure of income inequality. However, the MP-Gini is a conceptually superior measure.[6]

REFERENCES

S. Danziger, R. Haveman, and E. Smolensky, "The Measurement and Trend of Inequality: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 505–12.

W. Johnson, "The Measurement and Trend of Inequality: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 502–04.

C. J. Kurien, "The Measurement and Trend of Inequality: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 517–19.

J. J. Minarik, "The Measurement and Trend of Inequality: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 513–16.

J. N. Morgan, "The Anatomy of Income Distribution," *Rev. Econ. Statist.*, Aug. 1962, *44*, 270–83.

E. Nelson, "The Measurement and Trend of Inequality: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 497–501.

M. Paglin, "The Measurement and Trend of Inequality: A Basic Revision," *Amer. Econ. Rev.*, Sept. 1975, *65*, 598–609.

______, "The Measurement and Trend of Inequality: Reply," *Amer. Econ. Rev.*, June 1977, *67*, 520–31.

______, "The Measurement of Inequality: Reply," *Amer. Econ. Rev.*, Sept. 1979, *69*, 673–77.

---

[6] In a personal note, Paglin comments: "The few longitudinal income studies show the Gini of actual lifetime incomes to be about one-third less than Gini coefficients based on annual income data. The P-Gini approximates the lifetime Gini quite well, and this is what it is intended to do. The MP-Gini abstracts from the shape of the average age-income profile; the P-Gini does not. I believe this gives the P-Gini a useful attribute." Also see Paglin's 1979 "Reply."

Conglomerate Power without Market Power: The Effects of Conglomeration on a Risk-Averse Quantity-Adjusting Firm

By RALPH M. BRADBURD*

One of the interesting theoretical questions relating to conglomerate firms is whether conglomeration alone, unaccompanied by market power, can have any impact on firm behavior and market performance.

Many of those who argue that conglomeration without market power is behaviorally neutral obtain their results by assuming, implicitly or explicitly, that the firm maximizes profits (see Jesse Markham; Richard Miller). Their arguments are quite straightforward: the absence of market power precludes profit-maximizing behavior based on predatory pricing, mutual forbearance, advertising advantages, etc.; and if one assumes a *pure* conglomerate, among whose subsidiaries' products there are no substitution or complementarity relations,[1] then the necessary condition for maximizing overall conglomerate profits is to maximize each subsidiary's profits by setting its output to the level at which marginal cost equals price—exactly what each subsidiary would have done if it were an independent purely competitive profit-maximizing firm. Thus, if one assumes that the firm maximizes total profits, conglomeration per se will be behaviorally neutral in the absence of monopoly power.

Although the above argument is correct within the restrictions of its assumptions, our basic question is still not satisfactorily answered. Pure profit maximization is a very restrictive assumption, effectively implying—among other things—that the firm either operates in an environment in which there is no uncertainty, or that its management is indifferent to risk.[2] Such an assumption seems especially restrictive and inappropriate for modeling the behavior of conglomerate firms, since these firms have claimed risk reduction to be a major goal of their merger activity (see Harold Geneen).

The model developed below compares the optimal output of a risk-averse single product firm subject to demand uncertainty to that of a pure conglomerate subsidiary operating in the same market, assuming that the single product firm and conglomerate subsidiary are identically situated within that market[3] (which I shall call the *i*th product market) and that their managements are equally risk averse. The results indicate that, except in very special circumstances, conglomeration will not be behaviorally neutral even if all subsidiaries possess no monopoly power at all.

I. The Single Product Firm

Assume the existence of a single product purely competitive firm whose management is unable to predict with complete accuracy the market price at which the firm's output will be sold. Assume further that the firm's management is risk averse, that it maximizes the expected utility of the firm's income, and that it does so by maximizing the certainty equivalent of dollar profits π_i^*, which can be written as[4]

---

*Assistant professor of economics, Williams College. I would like to thank A. Mead Over, Jr., William Starnes, Gordon Winston, and an anonymous referee for their helpful suggestions.

[1] Thus, a change in the output of any one subsidiary will have no direct effect on the profit of any other subsidiary. This allows us to restrict our attention to conglomerateness per se.

[2] A risk-averse firm operating under conditions of uncertainty might also maximize expected profits if it believed that its actions had no effect on its level of risk exposure (see equation (3) below). This is an unlikely situation.

[3] In terms of size, expected sales, and other factors determining sales and profitability.

$$\pi_i^* = E(\pi_i) - \alpha\sigma^2_{\pi_i} \tag{1}$$

where $E(\pi_i)$ = the expected level of profit, $\sigma^2_{\pi i}$ = the variance of profit, and α = one-half the Arrow-Pratt measure of absolute risk aversion $[-U''(\$)/U'(\$)]$, and represents the dollar premium the decision maker would just be willing to pay to avoid an additional dollar of profit variance.

Let us assume that the purely competitive firm's sole short-run decision variable is its level of output q_i; that the firm chooses its optimal (expected utility-maximizing) output level in advance of production; and that it cannot alter its output level during the production period.[5] The optimal output level q_i^* is that at which

$$\frac{\partial\pi_i^*}{\partial q_i} = \frac{\partial E(\pi_i)}{\partial q_i} - \alpha\frac{\partial\sigma^2_{\pi_i}}{\partial q_i} = 0 \tag{2}$$

Inspection of (2) reveals that $\partial E(\pi_i)/\partial q_i$ must be positive (negative) at q_i^* if $\partial\sigma^2_{\pi i}/\partial q_i$ is positive (negative),[6] and assuming (to satisfy second-order conditions) that $E(\pi_i(q_i))$ and $\pi^*(q_i)$ are concave functions at their respective maxima, it follows that

$$q_i^* \lesseqgtr q_\pi \text{ as } \frac{\partial\sigma^2_{\pi_i}}{\partial q_i} \gtreqless 0 \tag{3}$$

where q_π = the expected-profit-maximizing level of q_i. To determine the sign of $\partial\sigma^2_{\pi i}/\partial q_i$, we must specify the form of the firm's profit function and the nature of the uncertainty to which the firm is subject.

The single product firm's dollar profit may be written as

$$P_i q_i - C_i q_i \tag{4}$$

where P_i = the price of the ith product and C_i = the level of average total cost.[7]

If we assume that the actual price of the ith product in any given period is equal to some expected price, $E(P_i)$, plus a normally distributed random term, ε, with $E(\varepsilon)=0$, then

$$\sigma^2_\pi = q_i^2\sigma_P^2 = q_i^2\sigma_\varepsilon^2 \tag{5}$$

and it follows that $\partial\sigma^2_{\pi i}/\partial q_i (=2q_i\sigma^2_\varepsilon)$ is positive. Combining this result with that of (3), we derive the familiar result that the output of a risk-averse competitive firm subject to demand uncertainty (of the variety indicated above) will be less than the output that would maximize expected profit.[8]

II. The Conglomerate Firm

In this section I lay the groundwork for comparing the optimal output levels of the conglomerate subsidiary and single product firm. As indicated in the introduction, it is assumed that the single product firm and conglomerate subsidiary are identically situated within the ith product market, that a change in the output of any given subsidiary has no direct effect on any other subsidiary's profit, and that the managements of the conglomerate subsidiary and single product firm are equally risk averse. It is further assumed that the conglomerate's central management participates in the output decisions of the individual subsidiaries, so that decisions regarding each subsidiary's output reflect that subsidiary's potential contribution to the conglomerate's

---

[4] It can be shown that the maximand of an expected utility-maximizing firm can be reduced to (1) when absolute risk aversion is constant over the relevant decision range, end-of-period wealth is equal to (certain and constant) initial wealth plus uncertain profits, and the distribution of uncertain outcomes is normal. See John Lintner, pp. 238–45.

[5] The production period may be defined simply as an interval of time over which the firm's output level is fixed.

[6] Note that α is positive by definition for a risk-averse decision maker.

[7] We assume that C_i is a determinate function of q_i and that $\partial C_i/\partial q_i$ is positive.

[8] See Ira Horowitz (chs. 8, 12, 13) for a very lucid exposition of several alternative models of the firm under uncertainty which arrive at this same result.

overall profit/risk position. Finally, to facilitate comparison with the single product firm, assume that when the conglomerate management determines the *i*th subsidiary's optimal output, it views the output levels of all other subsidiaries as given.

Like the single product firm, the conglomerate management attempts to maximize the expected utility of the firm's income, and must therefore maximize the certainty equivalent of the conglomerate's overall profits, or maximimize

$$(6) \qquad \pi_D^* = E(\pi_D) - \alpha\sigma_{\pi_D}^2$$

where π_D^* = the certainty equivalent of the conglomerate's total dollar profit, $E(\pi_D)$ = the conglomerate's total expected profit, and $\sigma_{\pi D}^2$ = the variance of total conglomerate profit.

When determining the optimal output level of the *i*th subsidiary, the conglomerate is assumed to view the output levels of the other subsidiaries as given. Thus, q_i is the sole decision variable, and must be set to the level which maximizes π_D^*. The optimal output level of the *i*th subsidiary q_D^* is therefore that at which

$$(7) \qquad \frac{\partial \pi_D^*}{\partial q_i} = \frac{\partial E(\pi_D)}{\partial q_i} - \alpha \frac{\partial \sigma_{\pi_D}^2}{\partial q_i} = 0$$

We will derive the relative levels of the single product and conglomerate firms' optimal output levels from differences in their first-order conditions for maximizing expected utility. But before we can do so, we must specify more completely the terms which enter into the conglomerate's certainty equivalent of profit.

The conglomerate's total dollar profit π_D is by definition the sum of all the individual subsidiaries' profits, or

$$(8) \qquad \pi_D = \sum_{j=1}^{n} \pi_j$$

where π_j = the *j*th subsidiary's profit. Because we are focusing on the *i*th subsidiary, we rewrite (8) as

$$(9) \qquad \pi_D = \pi_i + \pi_C$$

where π_i = the *i*th subsidiary's profit,[9] and π_C = the combined profits of all subsidiaries other than the *i*th.

If π_i and π_C are both subject to random variation, the variance of total conglomerate profits may be written as

$$(10) \qquad \sigma_{\pi_D}^2 = \sigma_{\pi_i}^2 + \sigma_{\pi_c}^2 + 2\operatorname{cov}(\pi_i, \pi_c) = \sigma_{\pi_i}^2 + \sigma_{\pi_c}^2 + 2\rho_{ic}\sigma_{\pi_i}\sigma_{\pi_c}$$

where $\sigma_{\pi D}^2$ = the variance of total conglomerate profits, $\sigma_{\pi c}^2$ = the variance of π_c, $\sigma_{\pi i}^2$ = the variance of the *i*th subsidiary's profit, $\operatorname{cov}(\pi_i, \pi_c)$ = the covariance of π_i and π_c, and ρ_{ic} = the coefficient of correlation between π_i and π_c.

It will be useful to digress momentarily to discuss the meaning of ρ_{ic}. Given our assumption that no substitution or complementarity relations exist between the conglomerate subsidiaries' products, and thus that changes in q_i have no causal effect on any π_j, it might appear that ρ_{ic} should be zero. However, if we view the profits of the various subsidiaries as being powerfully affected in either a similar or dissimilar fashion by exogenous forces—the fluctuation of the business cycle being the prime example—it becomes apparent that ρ_{ic} can readily assume a nonzero value. Thus, the profits of caviar-producing subsidiary *i* and the fine-watch-producing subsidiary *j* may be positively correlated, because both produce superior goods, even though a change in the price or output of the one will have no effect on the profits of the other. In principle, ρ_{ic} is easily obtainable. Assuming that past information provides a reasonable guide to the immediate future, one can obtain its value from time-series data on the individual subsidiaries' profits. Even if not publicly available, this information is accessible to the firm's management.

III. Comparing the Firms' First-Order Conditions

I have explicitly assumed that a change in the output of the *i*th subsidiary will have no effect on the expected profit of any other subsidiary, so that $\partial E(\pi_c)/\partial q_i=0$. It has also been assumed that within the confines of the ith market, the single product firm and conglomerate subsidiary are identically situated. It follows that

---

[9] Because $\pi_i(q_i)$ is identical for both the single product fim and conglomerate subsidiary, we use the same term π_i to denote either firm's profit.

$$\frac{\partial E(\pi_D)}{\partial q_i}=\frac{\partial E(\pi_i)}{\partial q_i} \tag{11}$$

Substituting (11) into (7), we may rewrite the conglomerate's first-order conditions as

$$\frac{\partial \pi_D^*}{\partial q_i}=\frac{\partial E(\pi_i)}{\partial q_i}-\alpha\frac{\partial \sigma^2_{\pi_D}}{\partial q_i}=0 \tag{12}$$

A comparison of (12) and (7) shows that any differences between the firms' first-order conditions with respect to q_i—and therefore between their respective optimal levels of q_i—must derive from differences in the effects of changes in q_i on their levels of profit variance. (Note that managerial risk aversion is equal by assumption.)

Taking the partial derivative of $\sigma^2_{\pi D}$ with respect to q_i, we find[10]

$$\frac{\partial \sigma^2_{\pi_D}}{\partial q_i}=\frac{\partial \sigma^2_{\pi_i}}{\partial q_i}+2\rho_{ic}\sigma_{\pi_c}\frac{\partial \sigma_{\pi_i}}{\partial q_i} \tag{13}$$

Our assumptions concerning the nature of the demand uncertainty faced by the firms ensure that $\partial\sigma^2_{\pi i}/\partial q_i$ and thus $\partial\sigma_{\pi i}/\partial q_i$ are positive. It follows that

$$\frac{\partial \sigma^2_{\pi_D}}{\partial q_i}\gtreqless\frac{\partial \sigma^2_{\pi_i}}{\partial q_i} \quad \text{as} \quad \rho_{ic}\gtreqless 0 \tag{14}$$

Thus, increases in q_i will cause the conglomerate firm's profit variance to increase by *more* than the single product firm's if π_i and π_c are positively correlated, and to increase by less if π_i and π_c are *negatively* correlated.

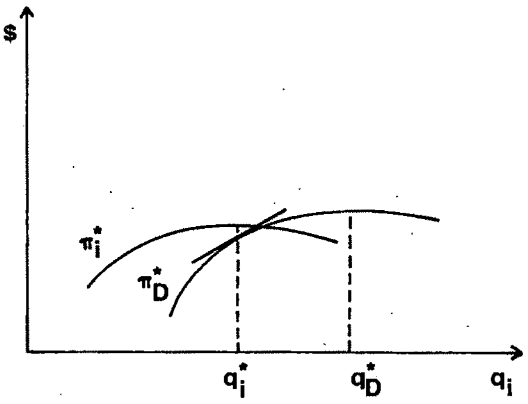


FIGURE 1. OPTIMAL OUTPUT LEVELS
The derivative of the conglomerate's maximand at the single product firm's optimal output level determines the relative levels of the firms' optimal output levels.

IV. Deriving the Relative Values of the Firms' Optimal Prices from Differences in their First-Order Conditions

Stability conditions require that π_D^* be a concave function at its maximum. It follows that[11]

$$q_D^*\gtreqless q_i^* \quad \text{as} \quad \frac{\partial \pi_D^*}{\partial q_i}\gtreqless 0 \quad \text{at } q_i^* \tag{15}$$

(See Figure 1 for the case of $q_D^*>q_i^*$.)

We know that at q_i^*,

$$\frac{\partial \pi_i^*}{\partial q_i}=\frac{\partial E(\pi_i)}{\partial q_i}-\alpha\frac{\partial \sigma^2_{\pi_i}}{\partial q_i}=0 \tag{16}$$

If $\dfrac{\partial \sigma^2_{\pi_D}}{\partial q_i}\gtreqless\dfrac{\partial \sigma^2_{\pi_i}}{\partial q_i}$

then[12] $\dfrac{\partial \pi_D^*}{\partial q_i}=\dfrac{\partial E(\pi_i)}{\partial q_i}-\alpha\dfrac{\partial \sigma^2_{\pi_D}}{\partial q_i}\lesseqgtr 0$ at q_i^*

and it follows from (15) that

$$q_D^*\gtreqless q_i^* \quad \text{as} \quad \frac{\partial \sigma^2_{\pi_D}}{\partial q_i}\lesseqgtr\frac{\partial \sigma^2_{\pi_i}}{\partial q_i} \tag{17}$$

---

[10] Because changes in q_i have no direct effect on any other subsidiary's profits, $\partial\sigma^2_{\pi c}/\partial q_i$ also equals zero.

[11] It is assumed that the local maximum of $\pi_D^*(q_i)$ in the neighborhood of $\pi^*(q_i^*)$ is the global maximum.

[12] See equation (12) and its derivation.

Combining this result with (14), we find that

$$q_D^* \gtreqless q_i^* \qquad \text{as } \rho_{ic} \lesseqgtr 0 \tag{18}$$

That is, assuming that increasing output will cause profit variance to increase, the optimal output level of the conglomerate subsidiary will be *greater* than that of the single product firm if π_i, the ith subsidiary's profits, are *negatively* correlated with the combined profits of the conglomerate's other subsidiaries, π_c; and *less* than that of the single product firm if π_i and π_c are *positively* correlated.

This result has a simple intuitive explanation. To maximize their respective certainty equivalents of profit, both firms adjust their output of q_i until the marginal cost of reduced expected profit equals the marginal benefit of reduced profit variance. Our assumptions that the firms are identically situated within the ith market and that their managements are equally risk averse ensure that both firms experience equal reductions in expected profit as q_i is reduced from the (expected) profit-maximizing level,[13] and also that both would place equal dollar value on any given marginal reduction in profit variance.

If ρ_{ic} is positive (negative), any given reduction in q_i will cause the conglomerate's profit variance to fall by more (less) than that of the single product firm, and its optimal reduction in q_i from the profit-maximizing level is therefore larger (smaller). Within the framework of our model, the optimal output level of the conglomerate subsidiary will only coincide with that of the single product firm if the coefficient of correlation between the profits of the ith subsidiary and the combined profits of the conglomerate's other subsidiaries is exactly equal to zero.[14]

V. Conclusions

The model developed in this essay compares the optimal output level of a risk-averse single product competitive firm facing uncertain demand to that of a conglomerate identical in every respect save that the latter is a subsidiary of a "parent" conglomerate. The results of the model demonstrate that if firms' managements are risk averse, and if the output decisions for each conglomerate subsidiary reflect that subsidiary's potential contribution to the conglomerate's overall profit/risk position, then in general, conglomeration will not be behaviorally neutral even when all subsidiaries possess no monopoly power at all.

REFERENCES

R. M. Bradburd, "Optimal Pricing by a Conglomerate Firm Under Conditions of Uncertainty," unpublished doctoral dissertation, Columbia Univ. 1976.

H. Geneen, "Conglomerates: A Businessman's View," in *Conglomerate Mergers and Acquisitions: Opinion and Analysis, St. John's Law Rev.*, Suppl. 1970, *44*, 723–42.

Ira Horowitz, *Decision Making and the Theory of the Firm*, New York 1970.

J. Lintner, "The Impact of Uncertainty on the 'Traditional' Theory of the Firm: Price Setting and Tax Shifting," in Jesse Markham and Gustav Papanek, eds., *Industrial Organization and Economic Development*, Boston 1970, 238–65.

Jesse Markham, *Conglomerate Enterprise and Public Policy*, Boston 1973.

R. Miller, "Conglomerate Mergers: A Monopoly Problem?," in *Conglomerate Mergers and Acquisitions: Opinion and Analysis, St. John's Law Rev.*, Suppl. 1970, *44*, 211–34.

---

[13] I have assumed that $\partial\sigma^2_{\pi i}/\partial q_i$ and $\partial\sigma^2_{\pi D}/\partial qi$ are positive, and that both firms must therefore reduce q_i below q_π to maximize their respective certainty equivalents of profits.

[14] If the level of managerial risk aversion of the single product firm is α_i while that of the conglomerate subsidiary is α_D, it can be shown that

$$q_D^* \lesseqgtr q_i^* \quad \text{as} \quad \rho_{ic} \gtreqless \left(\frac{\alpha_i}{\alpha_D}-1\right)\frac{\sigma_{\pi_i}}{\sigma_{\pi_c}}$$

For a more complete treatment, see my dissertation.

A Note on the Representation of Preferences in the Lindahl-Johansen Diagram

By GEOFFREY PHILPOTTS*

While much of the popularity of the Lindahl diagram undoubtedly arises from its ability to represent some rather complex relationships within a simple two-dimensional framework, the diagram and its extensions would appear to be somewhat prone to representational problems. Leif Johansen, for example, has questioned the appropriateness of an interpretation offered by Richard Musgrave (p. 76) whereby the familiar intersecting curves are taken to indicate the percentage of the cost of a public good that an individual would be willing to pay at various output levels, in contrast to the output preferred at various cost-share levels. In questioning the validity of such an interpretation, however, Johansen would also appear to have added some complications of his own by suggesting that this willing-to-pay concept would "...correspond to that indifference curve of person *A* which passes through the origin of the diagram (if such a curve does meaningfully exist)" (p. 350). Certainly a reflection of this idea in a recent reproduction of the diagram by Dennis Mueller (p. 400) does seem to create some difficulties. The object of this note, therefore, is to attempt to further clarify and extend some of the properties associated with the diagram.

While it is not fully clear as to why the intersecting curves were given such a prominent position within the apparently general bargaining context of Lindahl's original political decision model, a somewhat stronger case for their role in the determination of public decisions can perhaps be made within the more restricted environment suggested as a possibility by Johansen, in which annual decisions concerning public expenditures and the amount of taxes are envisioned as being agreed to by parties *A* and *B* subject to an existing (longer-term) tax structure, rather than both the level and distribution of taxes being simultaneously determined. Within this framework, the D_A and D_B curves of Figure 1 can indicate the most preferred public output levels of the participants at given tax-share rates, and the lower output (leftward) segments of these curves could then indicate the solution possibilities that would be attainable before mutual agreement breaks down between *A* and *B* as possible increases in expenditure are considered. Any specific solutions in this model would be dependent on the nature of the existing tax structure, and this can be assumed to somehow reflect and be responsive to the power of the participants. One solution that naturally invites special attention is the configuration of output and tax-share levels associated with the point of intersection *L*, which possesses the interesting property of enabling the output desires of both individuals to be simultaneously satisfied at the given tax-share rates. Few analysts, however, would appear prepared to follow Lindahl in his ascribing ethical significance to such a solution provided that an initial "just private distribution" had been previously attained. On the other hand, there may exist some temptation to occasionally assign rather too much significance to the intersection point regardless of the initial (or final) distribution, and there would not in general appear grounds for believing that a given "Lindahl-type" solution *L* would necessarily possess particularly desirable total welfare attributes, or be a particularly likely result in a general political bargaining context, or even be of great significance in resolving interparty conflict in view of the considerable leeway that remains for possible differences of opinion over the appropriateness of the tax-share levels themselves.

---

*University of Calgary.

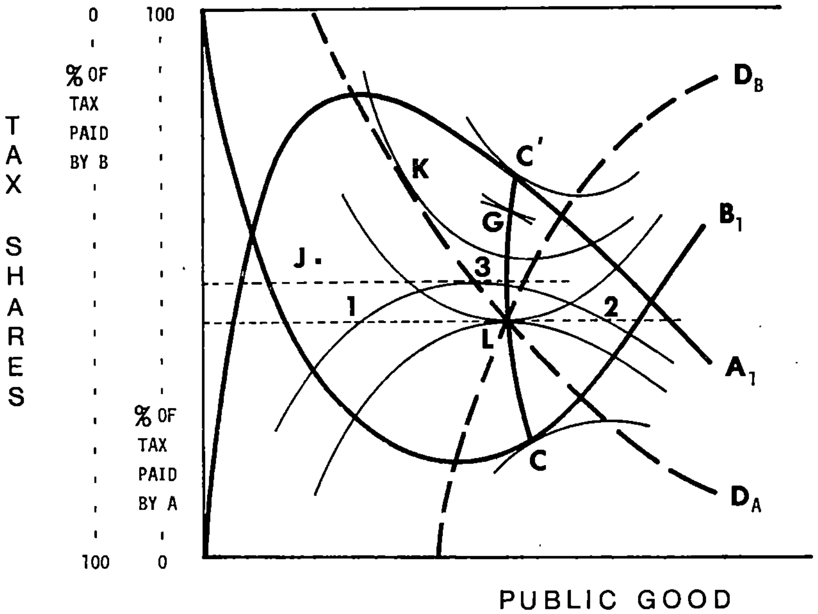


FIGURE 1

A considerable strengthening of the analytical properties of the simple intersecting curve diagram can of course be attained by explicitly introducing further preference information into it in the form of indifference curves. Figure 2 reproduces a Johansen-type source diagram for individual A, with a linear transformation function being assumed between private and public goods as indicated by the line $R_A Y$, and with OR_A representing A's initial endowment of the private good. As A's cost-shares decrease, which is indicated in this diagram by the pivoting price rays $R_A E$ and $R_A Z$, a share-consumption curve D_A' can be generated that is equivalent to the Lindahl D_A curve in Figure 1. In addition, by noting where the indifference curves intersect the various price rays (such as the points 1 and 2 on the Lindahl tax-share ray $R_A Z$, and the point 3 on the equal tax-share ray $R_A E$), it is possible to map these points onto the appropriate horizontal tax-share lines in Figure 1 and thereby generate the augmented Lindahl-Johansen diagram.

Within this framework the points of tangency (as represented here by the truncated segment $C'C$) between A's indifference curves and a similar but inverted set for individual B can then indicate the set of *Pareto-optimal* configurations whose significance has been emphasized in the general public expenditure analyses of Paul Samuelson. While this representation indicates that a Lindahl-type solution will be a member of this Pareto optimal group, it also serves to illustrate that any of the other members of this set would also appear capable of simultaneously satisfying both participants as regards the level of public output if the appropriate marginal costs (expressed here in terms of changing tax-share rates) were to be imposed, or at least accepted by A and B, and that any of these Pareto optimal solutions would also appear capable of representing the most socially desirable configuration (in terms of realized utility levels), regardless of the nature of any particular analytical pricing considerations that might be applied to this nonmarket environment.

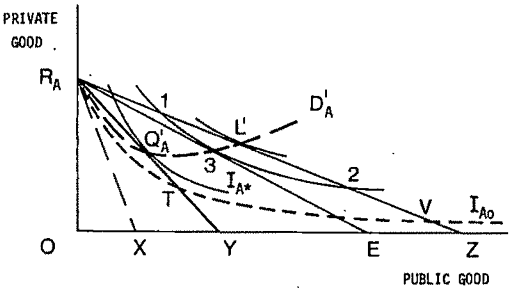


FIGURE 2

As further indicated by Samuelson (p. 105) and others, the coincidence of a Lindahl-type solution with a social optimum could be attained, if this were in fact desired (and if such a social optimum were known), by the appropriate relocation of the Lindahl-type solution through lump sum redistributions of the private good, although this property is perhaps not reflected clearly within this diagrammatic framework.

Of course, a great variety of alternate institutional arrangements and assumptions can be readily explored within these bounds, as for example the type of voting models recently discussed by Dennis Mueller. One possibility along these lines, that creates a version of the prisoner's dilemma from a modification of the Lindahl analysis, can be produced by introducing the assumptions that *A* and *B* indicate their rankings of the possible output levels at given tax-share rates (possibly through some voting mechanism) and agree to proceed only if they indicate agreement as to the best output level. If both of them were honest then they should produce the solution *L*, but if one or both tried to secure some free-rider advantage by understating their true preferences, then at least one and possibly both could be made worse off than they would have been in the jointly honest situation, (perhaps ending at *J*, or possibly at *K* if *A* were honest and *B* were well informed and optimally dishonest).

Another example of a rather different type of decision procedure involves a more arbitrary iterative movement from the initial situation with no public good to some final solution by a series of steps that are each indicated to be mutually satisfactory by *A* and *B*, and which are proposed by an impartial observer. While the ultimate solution to such a process would tend to be somewhat indeterminate, it would appear restrictable to Pareto optimal points if both individuals were honest, or at least honest in the final steps, and in any event it should not be Pareto inferior to the initial private goods position whether they were honest or not. One characterization of this set of Pareto-preferred solutions, which would appear to be in accord with Johansen's comments on the willing-to-pay concept, is produced by taking indifference curves similar to A_1 and B_1 in Figure 1 as corresponding to the utility levels of individuals *A* and *B* that would prevail in the absence of any public good. The eye formed by the intersection of these curves would then indicate the set of points that are Pareto preferred to the initial situation, and the line $C'C$ would correspond to the locus of Pareto optimal points within this set.

The problem with this representation, however, is that under conventional assumptions if any indifference curve for *A* were to attain a maximum (or a minimum for *B*) at some positive level of the public good in the Lindahl-Johansen diagram, then it could never intersect the vertical axis; or in terms of Johansen's source diagram (Figure 2) given the usual convexity assumptions, if any indifference curve attains a tangency on some price line in the interior of this space, then it could not pass through individual *A*'s initial private distribution point R_A.

While this consideration would invalidate indifference curves with the properties exhibited by A_1 and B_1 in Figure 1, it is clear that some representation should be possible of the set of points that are Pareto preferred to a situation in which no public good is being provided, and the dashed lines A_o and B_o in Figure 3, would appear to be an appropriate pair of indifference curves for this task. To generate these curves the I_{A_o} indifference curve in the commodity space of Figure 2, which includes the no-public-production point R_A, is transformed into the

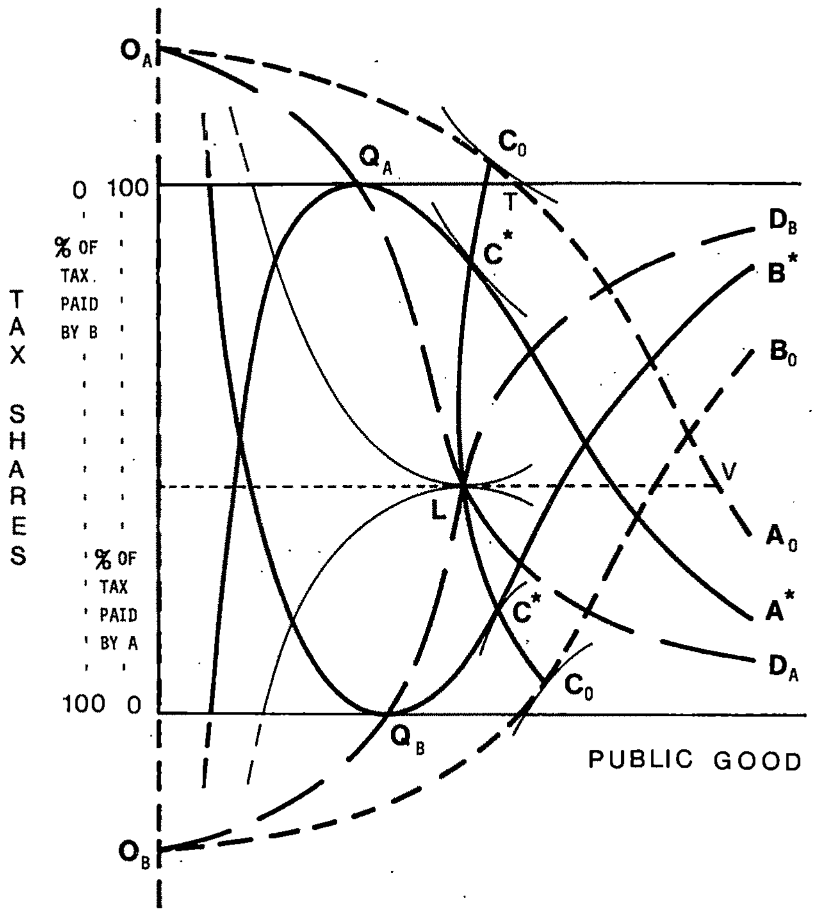


FIGURE 3

A_o curve in the public good/tax-share space of Figure 3, with, for example, the levels of the public good that are associated with the intersection of I_{A_o} and the tax-share lines $R_A Y$ and $R_A Z$ being indicated by T and V, respectively, in both diagrams. The maximum price that an individual could pay for some supply of the public good without incurring a welfare loss, may, at least for low levels of provision, actually exceed the cost of the public good, and such a price ($R_A X$ in Figure 2) for the first marginal unit yields a positioning of the point 0_A in Figure 3 at a tax share for A that is in excess of 100 percent; although situations in which 0_A could occur at less than 100 percent would be possible if A did not want any of the public good unless he could get it at a less than full-cost price. (If any social decision about the supply of the public good actually involved tax shares in excess of 100 or less than 0 percent, then this might be interpreted, if no budget imbalance were allowed, as a decision that also involves some redistribution of the private good between A and B.)

The full set of points on the A_o and B_o indifference curves can therefore be seen to indicate the maximum amounts of the public good that could be provided at given tax-share rates to which the respective individuals would be indifferent in comparison with the no-public-good situation (the vertical axis). These points would therefore also indicate the maximum tax burden that an individual would be willing to bear at any given level of public output if it were a choice between this level of output or no public output at all, and these curves could therefore be taken as a representation of the willing-to-pay concept. The individuals' most desired levels of the public good at any given tax-share rates would of course be indicated by points that lie inside these curves, and these would correspond to the points of tangency of the respective indifference curves with the horizontal tax-share lines, as shown by the dashed D_A and D_B loci.

Finally, it may be observed that while the truncated eye formed by the curves A_o and B_o would delineate the set of points that are Pareto preferred to a situation devoid of any public good, it would seem that such boundaries might not be of particular relevance to the participants in a noncoercive society. Rather, it could be expected that in the type of situation represented in Figure 3, where both participants would appear prepared to obtain some private supply of the public good at a full-cost price rather than go without, their interest would tend to center on those social agreements that would represent improvements over the best situations that they could attain on their own. The relevant indifference curves would then be A^* and B^*, which are tangent to the horizontal private opportunity lines for A and B at Q_A and Q_B, respectively, and these curves might be interpreted here as indicating the upper limit to the individuals' willingness to pay under some collective public provision arrangement, with the line C^*C^* now indicating the possible Pareto optimal solutions within these bounds. The exact configuration of taxes and public good output that might be attained in a given social situation will of course depend on the preferences and strategic behavior of the participants and the nature of the institutional environment in which they operate.

REFERENCES

L. Johansen, "Some Notes on the Lindahl Theory of Determination of Public Expenditures," *Int. Econ. Rev.*, Sept. 1963, *4*, 346–58.

E. Lindahl, "Just Taxation—A Positive Solution," in Richard Musgrave and Alan Peacock, eds., *Classics in the Theory of Public Finance*, New York 1962, 168–76.

D. Mueller, "Public Choice: A Survey," *J. Econ. Lit.*, June 1976, *14*, 395–433.

Richard Musgrave, *The Theory of Public Finance*, Toronto 1959.

P. Samuelson, "Pure Theory of Public Expenditure and Taxation," in Julius Margolis and Henri Guitton, eds., *Public Economics*, Toronto 1969, 98–123.

Measuring the Expected Real Rate of Interest—An Exploration of Macroeconomic Alternatives: Comment

By John J. Merrick, Jr.*

In an article recently published in this *Review*, J. Walter Elliott developed a structural approach to measuring the expected real rate of interest under the assumption that expectations are formed rationally in the sense of John Muth. Elliott's avowed strategy was, first, to search over alternative macro-economic models to identify the "relevant" theory for explaining real rate fluctuations, and second, to solve for this relevant theory's prediction of the real interest rate which, according to Muth, equals the market's expected real rate. In Elliott's heuristic characterization of the model-identification process, the relevant model is the one which forecasts the *ex post* real rate with the minimum prediction error. By combining two "structural" macro-economic models, neo-Keynesian and loanable funds, with two explanations of inflationary expectations, Elliott produces four specifications over which to "explore." The "best" measure of the expected real rate of interest is the one generated by the most predictive-efficient model of the four alternatives.

Elliott's work addresses the question of uncertainty over model selection under the Muthian premise that market agents judge alternative models on their efficiency in forecasting the variable of interest. Information is scarce, agents use it economically, and any correctable forecast biases are corrected. In the limit, the best model is the true model. However, though the Elliott paper starts from a firm Muthian base, it quickly develops an internal contradiction once the "exploration" process begins. The fundamental flaw is a failure to properly incorporate a second Muthian premise which requires that expectations be modelled as if the agents understood the economic process by which the phenomenon to be forecast evolved and based their predictions upon the distribution of outcomes as generated by the actual process. Thus, to ferret out the expression for "rationally formed" expectations of inflation one must look to the structure of the "true" macro-economic model for the process generating inflation. If the structural model properly specifies the inflation process, the rational expectations hypothesis will be sufficient to yield the reduced-form solution for expected inflation.

In Elliott's paper "rational" market agents are assumed to form their expectations on the basis of two alternative specifications for inflation: the monetary hypothesis and the wages hypothesis. These two theories of the inflation process are developed independently of the forces generating price level growth in either of Elliott's two structural models. The monetary hypothesis is based upon assumptions that the equilibrium price level is proportional to the nominal money stock, that actual prices adjust toward equilibrium through a partial adjustment mechanism, and that the money stock one period ahead is proportional to the currently observed money stock. Given these three assumptions, the expected value of the current inflation rate conditioned on information currently available is expressed by Elliott's equation (16) as a distributed lag on past money stocks.

Monetary Hypothesis

$$(16) \qquad {}_t\pi_t^* = \Phi \sum_{i=0}^{\infty} \mu_i m_{t-i} + e_{1,t}$$

Elliott's wages hypothesis is derived from a Walrasian-type labor market adjustment mechanism, an autoregressive model of money wage rate fluctuations, and a distributed-lag model of real wage movements that incorporates growth in the equilibrium real wage. The conditional expected value of the current inflation rate is then given by Elliott's (34).

---

*Graduate student, Brown University. I am grateful to Robert King, William Poole, the managing editor, and a referee for helpful comments.

Wages Hypothesis

$$(34)\quad {}_{t}\pi_{t}^{*} = \sum_{i=0}^{\infty} \gamma_i(\log W_{t-i} - \log W_{t-i-1}) + \beta[\log(W/P)_{t}^{e}] - \sum_{i=1}^{\infty} \psi_i[\log(W/P)_{t-i}^{e} - \log(W/P)_{t-i-1}]$$

where W, P, and $(W/P)^e$ are the levels of money wages, prices, and equilibrium real wages, respectively.

The point I would like to stress is that neither of Elliott's equations, (16) nor (34), is the general rational expectations solution for the expected rate of inflation given either of Elliott's structural models. This point is somewhat self-evident since no reference was made to either the neo-Keynesian or loanable funds models in the derivations of (16) and (34). In effect, Elliott's rational agents use expectations functions which are logically independent of those implied by his structural models under the rational expectations hypothesis. Since Elliott's inflationary expectations equations are not those generated by the two structural models in question, the expectations variables included in his expected real interest rate regressions are misspecified. Thus, Elliott's measures of the expected real rate of interest are not derived from rational expectations considerations.[1] In fact, Elliott himself provides the statistical evidence needed to reject the supposed "rationality" of his most efficient expected real rate measure. Equation (D) of his Table 4 shows that there is a systematic relationship between the current real rate prediction error and the previous period's *ex post* inflation rate. According to that equation, the lagged inflation rate explains 25 percent of the variation in the current real rate forecast error. Since the lagged inflation rate is assumed to be part of the information set at the time current real rate expectations are formed, Elliott's rational prediction measure violates the most fundamental precept of Muthian rationality: no easily correctable forecast bias can be allowed to persist.

Moreover, Elliott's dichotomy between expectations and structure causes problems in his neo-Keynesian/monetary model that extend beyond its lack of Muthian rationality. After adding his monetary equation for expected inflation (16) to the neo-Keynesian block to determine the expected real rate of interest, Elliott is left with a model that in the general case is internally inconsistent. Given values for the exogenous variables and the current state of expectations, the model implies a market-clearing equilibrium price level. Note that the monetary balance condition (Elliott's equation (5)) requires the difference of the logarithms of the money stock and the equilibrium price level to equal $c_1 y_t + c_2 r_t + v_{3,t}$. Yet, in deriving (16) Elliott assumed this difference to be constant over time. Therefore, unless $c_1 y_t + c_2 r_t + v_{3,t}$ was constant throughout his entire sample period—and there is no reason to suppose that this was the case—Elliott's model breaks down. While an internally consistent macro-economic model need not incorporate rational expectations, the inconsistency of the neo-Keynesian/monetary model is due directly to its lack of Muthian rationality.

In general, the problem of measuring the "rationally expected" real rate of interest using a structural approach is not as simple as Elliott's misuse of the rational expectations hypothesis makes it appear. Consider Elliott's neo-Keynesian model in this regard. The rational expectations solution for the current expected rate of inflation in this model depends crucially upon the agents' forecasts of *future* changes in the model's exogenous variables

---

[1] This dichotomy between the structure and the inflationary expectations function is even more curious when one considers that after forming predictions about the inflation rate without reference to the informational content of the structure, a structural equation is used to generate estimates of the market's expectation of the real rate of interest.

$$
\begin{aligned}
{}_{t}\pi_{t}^{*} = & \left(\frac{G_o}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\left[{}_{t}p_{t+i}^{*} - {}_{t-1}p_{t+i}^{*}\right] \\
& + \left[\frac{G_2}{1-G_o}\right]\sum_{i=0}^{\infty} h^{i}\left[{}_{t}m_{t+i+1}^{*} - {}_{t}m_{t+i}^{*}\right] \\
& + \left(\frac{G_3}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\left[{}_{t}Z_{t+i+1}^{*} - {}_{t}Z_{t+i}^{*}\right] \\
& + \left[\frac{G_4}{1-G_o}\right]\sum_{i=0}^{\infty} h^{i}\left[{}_{t}d_{t+i-1}^{*} - {}_{t}d_{t+i}^{*}\right] \\
& + \left(\frac{G_5}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\left[{}_{t}X_{t+i+1}^{*} - {}_{t}X_{t+i}^{*}\right]
\end{aligned}
$$

where ${}_{t}\pi_{t}^{*} \equiv E_{t}[p_{t+1} - p_{t}]$, the G_i and h are combinations of structural parameters from the reduced-form equilibrium price level solution ($h < 1$), X_t is a combination of structural parameters and the three stochastic terms, and the starred variables represent the market expectations of the variables defined by Elliott. The "forecast up-date" term $[{}_{t}p_{t+i}^{*} - {}_{t-1}p_{t+i}^{*}]$ can also be expressed as a linear combination of the revisions in the expectations of future exogenous variables resulting from the new information accruing in time t.[2]

Obviously, this type of solution does not yield itself very easily to empirical study. To make the model empirically operational, some assumptions must be made about the way in which the market predicts future values of the money supply, capacity output, government spending, and the three disturbance terms. The natural assumption would be to hold that the market also forms its expectations of these variables rationally, but this would require a complete characterization of the actual processes governing the future fluctuations of these variables.

One way to bridge the gap between the forward-looking solution and the current data base would be to specify the exogenous variables as conforming to pure autoregressive time-series processes.[3] By the chain rule of forecasting, the forward-looking solution for expected inflation could be expressed in terms of current and lagged values of the exogenous variables, the number of lags depending upon the order of the autoregressive process assumed to be driving each variable

$$
\begin{aligned}
{}_{t}\pi_{t}^{*} = & \sum_{\tau=0}^{k_1} \alpha_{1,\tau} m_{t-\tau} + \sum_{\tau=0}^{k_2} \alpha_{2,\tau} Z_{t-\tau} \\
& + \sum_{\tau=0}^{k_3} \alpha_{3,\tau} d_{t-\tau} + \sum_{\tau=0}^{k_4} \alpha_{4,\tau} X_{t-\tau}
\end{aligned}
$$

As the "proper" determinants of expected inflation these current and lagged exogenous variables must be included in the neo-Keynesian model's expected real interest rate regression. Elliott's original two-stage least squares (*2SLS*) approach would be "corrected" by adding the appropriate prespecified number of lagged exogenous variables to his original first-stage regression of real output on the current values of the exogenous variables to determine the proper

---

[2] Actually, the inflation forecasting problem in Elliott's model is really that of predicting the next period's price level since the market's information set is assumed to include the current value of the price level. In this case, the solution would involve only the expected stocks of the exogenous variables and the current price level and would not include the forecast update term $({}_{t}p_{t+i}^{*} - {}_{t-1}p_{t+i}^{*})$,

$$
\begin{aligned}
{}_{t}\pi_{t}^{*} = \underset{t}{E}(p_{t+1}) - p_{t} = & \left(\frac{G_2}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\,{}_{t}m_{t+i+1}^{*} \\
& + \left(\frac{G_3}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\,{}_{t}Z_{t+i+1}^{*} + \left(\frac{G_4}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\,{}_{t}d_{t+i+1}^{*} \\
& + \left(\frac{G_5}{1-G_o}\right)\sum_{i=0}^{\infty} h^{i}\,{}_{t}X_{t+i+1}^{*} - p_{t}
\end{aligned}
$$

For a discussion of the nature of "forward-looking" solutions and the terminal condition imposed see Thomas Sargent and Neil Wallace.

[3] Robert Barro utilized a specific structural model of the money supply rule to derive a "prediction" equation for the monetary growth rate to generate a measure of "anticipated" money growth. In a forward-looking context, however, such a structural approach would only serve to expand the present model unless forecasts of all of the "new" forcing variables could be tied to the current data base.

$\hat{y}_t$. It is crucial to note that the appropriate number of lags for each variable in the $\hat{y}_t$ equation should be determined independently since the lagged values enter the model only through their use in predicting their own future values.[4,5]

The simple result that rational inflationary expectations depend on present and past values of the exogenous variables comes about because of the autoregressive rules used to forecast future values of those variables. This in no way lends support to using such schemes in a naive fashion. To highlight this point, consider the case in which the Federal Reserve announces that henceforth it will maintain a specified constant monetary growth rate. If market participants believe that the Fed will carry out its announced intentions, they would be "irrational" to continue forecasting the money supply according to the outdated autoregressive formula. Inflationary expectations, which depend on monetary growth expectations, would also be altered. Past money stock information would then be irrelevant to the forecasting process.[6]

By coincidence, *if* one is willing to make the autoregressive assumptions discussed above, Elliott's monetary equation (16) would take the same form as the "proper" rational expectations solution for the expected inflation rate for the neo-Keynesian model. Yet, even in this special case, Elliott's inflationary expectations function would constrain the coefficients on all of the nonmonetary variables to equal zero.[7] However, in light of this coincidence, one can readily understand why Elliott's neo-Keynesian/monetary model outperforms the neo-Keynesian/wages variant in predicting the *ex post* real interest rate. This occurs because the monetary model is a distant cousin to the true model, whereas the wages model is not a family member. As such, the monetary model cannot be statistically bested by the wages model under the joint null hypothesis of rational expectations and the "true" neo-Keynesian structure.[8] Thus, Elliott's comparison of the prediction results of the two models is not statistically meaningful.[9]

---

[4] Although Elliott chose to use *2SLS* to generate the expected real rate series from his aggregate demand schedule, the series could be generated directly from the reduced-form representation. With the appropriate grouping of parameters the equation would take the following form

$$\rho_t = \sum_{\tau=0}^{k_1} \beta_{1,\tau} m_{t-\tau} + \sum_{\tau=0}^{k_2} \beta_{2,\tau} Z_{t-\tau} + \sum_{\tau=0}^{k_3} \beta_{3,\tau} d_{t-\tau} + \sum_{\tau=0}^{k_4} \beta_{4,\tau} X_{t-\tau}$$

where the *ex post* real rate is substituted for ρ_t and the predicted values $\hat{\rho}_t$ are interpreted as expected real rates.

[5] One significant problem to be faced is that the $X_{t-\tau}$ $(\tau=0,\dots,k_4)$, the linear combination of disturbance terms, are unobservable and must be excluded from the regression. The exclusion of relevant variables from a regression is a standard econometric problem. The estimated coefficients will be biased to the extent that the excluded and included variables are correlated. For present purposes, the exclusion of the $X_{t-\tau}$ yields no immediate problem on this account. The difficulty occurs if the $X_{t-\tau}$ are *not* correlated with the included variables. In this case, the contribution of the $X_{t-\tau}$ would be picked up solely by the residual. As is most clearly shown by the reduced-form representation in fn. 4, the exclusion of the $X_{t-\tau}$ would complicate the overall problem of measuring the expected real rate. It is also of interest to note that at least one component of $X_{t-\tau}$, the supply-side disturbance $v_{1,t}$, must show substantial positive serial correlation to account for the "business cycle" since Elliott's neo-Keynesian specification does not allow for "persistence" effects.

[6] Within Elliott's rational framework, a credible Fed announcement of a new policy, effective a year from now, would *not* change current inflationary expectations. Expectations about the rate of inflation would continue to be formed through his equation (16) on the basis of past money stock data. Certainly, this is a somewhat inverted view of a rational expectations hypothesis (*REH*) world: new information is ignored and old information is incorporated. The source of this paradoxical result is Elliott's misuse of the *REH* in deriving (16).

[7] It is certainly plausible that the nonmonetary variables could fail to add significant explanatory power for the rate of inflation to the set of monetary variables. However, coefficient restrictions of this type are open to explicit statistical tests.

[8] One qualification should be mentioned. The wages model could possibly predict the inflation rate and real interest rate better than the monetary model if lagged wages forecast future values of the exogenous variables better than autoregressions.

[9] Of course, such a comparison *would* be meaningful if it showed that the wages model outperformed the

There is no rationality to Elliott's mixing and matching of expectations functions and structural models. The correct exploration strategy would have compared the real rate predictive efficiency of alternative structural models after having imposed the rational expectations hypothesis upon those models' reduced-form expressions for the current rate of inflation. In general, much stronger assumptions than those Elliott uses about how market agents forecast must be introduced to make such an exploration meaningful. This is not to state that the "true" rationally expected real interest rate series does not have the same nonconstant, noncyclical properties reported by Elliott; only that, as presented, his findings cannot be supported by claims of "rationality."

---

monetary model. This would lead to a rejection of the joint null hypothesis, if one could rule out the case discussed in fn. 8.

REFERENCES

R. J. Barro, "Unanticipated Money Growth and Unemployment in the United States," *Amer. Econ. Rev.*, Mar. 1977, *67*, 101–25.

J. W. Elliott, "Measuring the Expected Real Rate of Interest: An Exploration of Macroeconomic Alternatives," *Amer. Econ. Rev.*, June 1977, *67*, 429–43.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961, *29*, 315–35.

T. J. Sargent and N. Wallace, "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *J. Polit. Econ.*, Apr. 1975, *83*, 241–54.

R. J. Shiller, "Rational Expectations and the Dynamic Structure of Macroeconomic Models," *J. Monet. Econ.*, Jan. 1978, *4*, 1–44.

Measuring the Expected Real Rate of Interest— An Exploration of Macroeconomic Alternatives: Reply and Further Thoughts

By J. WALTER ELLIOTT*

John Merrick's critique of my 1977 study of real interest rate movements rests on two grounds. The criticism that occupies most of his attention is concerned with whether my model remains internally consistent when either of the two models of inflationary expectations are added to the macro-economic models of the real rate of interest.

Merrick finds that each of my macroeconomic models of the real rate has embedded in it a model of inflationary expectations when rational expectations are applied to it, so that my procedure of adding alternative models of inflationary expectations to each macro-economic structure produces a misspecification of the models and constitutes a fundamental flaw in the procedure used. Since this "fundamental flaw" contention is by far the most serious Merrick makes, I shall discuss it first.

The neo-Keynesian model given by expressions (3), (4), and (5) in my paper contains four endogenous variables: real income y; the price level π; the nominal interest rate r; and the real interest rate π. The model containing these variables has three equations and three exogenous variables. The result is an underidentified model that does not by itself lead to a solution for the endogenous variables or yield to reliable statistical estimation. What is needed to resolve this familiar problem is sufficient additional theoretical restrictions to identify the model. In my work, I developed two alternative ways to identify the model and thus complete the specification of the neo-Keynesian framework. These alternatives each add a new theoretical restriction to the original set. One of these is a monetary explanation of $r-\pi$ while the second is a labor market explanation. Merrick does not contest the question of whether the completed models are identified econometrically. Nor does he contest the theoretical motivation of either the monetary or wage explanation of $r-\pi$.

Then, where exactly does my model become "internally inconsistent" in Merrick's eyes? Judging from Merrick's adjectives, it is because my alternative explanations of $r-\pi$ do not follow the specific Lucas-Sargent-Wallace version of the rational expectations hypothesis that they developed in connection with the neutrality of money and the Phillips' curve. This was a deliberate choice on my part, but one that apparently needs some additional discussion.

In my paper, I explained the sense in which my expectations estimates were "rational" in three passages (all on p. 430): "Estimates of the real rate are formulated from Muthian premises...; John Muth's concept of such an expectation is that it is rational if it is equal to the prediction of relevant theory;... an expectations measure that reflects current information more fully than another measure is 'rational' in that it produces more accurate predictions than other alternatives."

In constructing the alternative models in my study, I used these ideas to support two assumptions. First, if as a testable alternative the public believed that actual prices were related predominantly to the money supply, then their expectation of future prices, $r-\pi$, would be as strongly and in the same structural way related to the money supply. The same reasoning process supported a second wage explanation of inflationary expectations. If the public associated price movements predominantly with labor market variables, then these variables would predominantly govern their expectations of future prices.

---

*University of Wisconsin-Milwaukee.

This approach is indeed formulated from Muthian premises. But, of more importance, it represents a tailoring of Muth's ideas to a particular concern of my study—that of examining alternative structural models of inflationary expectations. By formulating and using these alternative structural models, evidence is obtained as to whether money market or labor market information is more important in influencing inflationary expectations.

Under the usual Lucas-Sargent-Wallace approach, any variable that enters the reduced-form of the otherwise incomplete model becomes an argument in the added equation for $r-\pi$. Thus, my two alternative structural hypotheses that the public follows either monetary or labor market rules in formulating inflationary expectations are replaced by the single premise that all the exogenous and lagged endogenous variables are linearly combined by the public in forming its inflationary expectation. This is an entirely appropriate abstraction in the case of studies such as those of Lucas or Sargent-Wallace. But, in my study, this procedure would suppress and abstract from an important question that it was my purpose to examine. All of this illustrates that Muth's concept of rational expectations can be used in more than one way in empirical economic research.

Since Merrick rules out all but the version he develops of the Lucas-Sargent-Wallace approach to rational expectations, he finds my development of alternative structural models for $r-\pi$ inconsistent with "the" rational expectations hypothesis. My development clearly does not follow the Lucas-Sargent-Wallace approach. But it is fully consistent with the broader Muthian premise upon which it is founded, and it is fully appropriate for my purpose of exploring alternative macro-economic structures.

The second criticism Merrick finds with my method of analysis relates to the monetary balance equation in the neo-Keynesian structure (my expression (5)) and the monetary model of inflationary expectations (my expression (16)). Expression (5) states that, in equilibrium, the *log* of the real money stock $(m-p)$ is in balance with the *log* of the demand for real balances, where the latter is a positive function of real income and a negative function of the interest rate. In deriving expression (16) by using a monetary rule, my expression (10) asserts: $p_t^e = a + m_t$, where p_t^e is the *log* of an equilibrium price and m is the *log* of nominal money balances.

Merrick's contention is that since expression (10) fixes the value of $m_t - p_t^e$ to be constant through time, the right-hand side of my expression (5), given by $c_1 y_t + c_2 r_t + v_{3,t}$, is required to be constant through time which he contends is unlikely: thus he contends that my model breaks down. In this analysis, Merrick apparently overlooks the fact that the price that appears in the monetary balances expression is the actual price p; while the price appearing in expression (10) is an equilibrium price towards which the actual price adjusts according to the partial-adjustment mechanism described by expressions (11)–(13).

According to (5), monetary equilibrium is disturbed when changes occur in m, p, r, or y. Changes in p_t^e do not of themselves impact upon the monetary balance equation. Rather, the sequence goes like this. When m is increased, p_t^e adjust proportionately upward. Actual prices then start to change, eroding the real money stock and causing an increase in r. Finally, the increase in p_t and r_t is sufficient to reduce real money demand to the fallen supply. So, although the monetary model of inflationary expectations requires that $m_t = p_t^e$ be constant through time, the monetary balance equation does not require that $c_1 y_t + c_2 r_t + v_{3,t}$ be constant through time—only that it be set equal to a generally nonconstant $m_t = p_t$ at each point in time.

To conclude, I have taken two incompletely specified macro-economic models and completed each of them in two different ways. The result is four logically complete and statistically estimable models that each explain real income, prices, the nominal interest rate, and the real rate by differing theoretical premises. Merrick has not successfully shown there to be a problem with this research method. However, he has suggested two additional and highly general ways to complete the macro-economic structures—by assuming along the lines of Lucas-Sargent-Wallace that inflationary expectations are exclusively set by a linear combination of past values of the exogenous variables from each partial model. I have argued that my purpose suggested a specific structural approach to be preferable.

The two alternative representations of inflationary expectations I examined reflect points of view found prominently in the literature. But, there are more than Merrick's two unexplored alternatives. There are as many plausible ways to complete the macro-economic models in my paper as there are ideas on how inflationary expectations are set. In this or any other empirical study, the list of alternative hypotheses will generally not be complete so that empirical studies such as mine will reach less than global conclusions. Nonetheless, such results will be useful in understanding how the expectations of economic agents are most likely set.

REFERENCES

J. W. Elliott "Measuring the Real Rate of Interest: An Exploration of Macroeconomic Alternatives," *Amer. Econ. Rev.*, June 1977, *67*, 429–43.

J. J. Merrick, "Measuring the Real Rate of Interest—An Exploration of Macroeconomic Alternatives: Comment," *Amer. Econ. Rev.*, June 1980, *70*, 493–97.

T. J. Sargent and N. Wallace, "'Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *J. Polit. Econ.*, Apr. 1975, *83*, 241–54.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961, *29*, 315–35.

R. E. Lucas, Jr., "Some International Evidence on Output-Inflation Tradeoffs," *Amer. Econ. Rev.*, June 1973, *63*, 326–34.

Excess Burden, Benefit Taxation, and Efficiency in Public Expenditure

By THOMAS S. MCCALEB*

If direct user charges for public goods are infeasible, is there a system of taxation which might promote efficiency in public expenditure where the supply of public goods is determined by voting? This is the central issue addressed by the modern theory of benefit taxation, particularly as exposited by James Buchanan. He provided an affirmative answer to the question just posed and developed a formula for distributing the tax burden among the individual voters. The formula implies that an efficient level of public expenditure would be collectively chosen no matter what voting rule might have been established at the consitutional stage of decision making. Application of this neutrality formula, however, requires knowledge of the elasticities of demand for public goods with respect to the variables which influence that demand. It is natural, then, that the formula has reappeared in several recent studies which attempt empirical estimation of demand functions for various public goods. These studies propose a test based on Buchanan's formula for determining the extent to which the supply of a public good in a particular community deviates from the efficient quantity.

Both Buchanan's work and the empirical demand studies just noted. implicitly abstract from the excess burden of taxation. Indeed, most of the literature on the economics of public goods neglects the effects which excess burden might have. Two notable exceptions are papers by David Johnson and Mark Pauly and by Anthony Atkinson and Nicholas Stern. The Johnson-Pauly paper shows that in the presence of distortionary taxation the familiar marginal conditions first presented by Paul Samuelson no longer characterize a Pareto optimum. Their analysis, however, is confined to a voluntary exchange model, and they fail to extend the analysis to show how the existence of excess burden alters Buchanan's formula. Atkinson and Stern, expanding on an argument in Joseph Stiglitz and Partha Dasgupta, incorporated the excess burden of taxation into the marginal social benefit obtained from public expenditure. They show that, absent complementarity between a public good and the private goods which are taxed to pay for it and provided the taxed private goods are not inferior, the socially optimal quantity of the public good will be reduced if distortionary taxation must be used. Their analysis is confined to a planning model rather than a voting model, however, so that the voting rule and the distribution of the tax burden among the individual voters are of no interest to them.

The objective of this paper is to develop the implications of distortionary taxation for Buchanan's neutrality formula. A new formula which takes account of excess burden is derived along with a sufficient condition for equivalence between this modified criterion and the original formula. The implications of this adjustment for the efficiency test proposed in the empirical demand studies are then investigated. In particular, it will be shown that use of the original formula to determine the efficient level of public expenditure leads to quantitatively erroneous, and in some cases, qualitatively erroneous, conclusions. Some crude estimates of the magnitude of the potential error are also provided.

---

*Assistant professor of economics, Rice University. This investigation was supported by University of Kansas General Research grant allocation No. 3673-X038. I acknowledge the helpful comments of Ronald S. Warren on an earlier draft.

I. Excess Burden and Neutrality of the Voting Rule

Assume that all individuals are alike in their preferences for some public good z, but that they differ from one another in terms of some variable y which together with the tax price imposed on them determines their individually preferred quantities of the public good. The individual demand functions can therefore be written as

$$z_i = f(t_i, y_i) \qquad i = 1, \ldots, r \tag{1}$$

where t denotes tax price and the subscript refers to the ith individual.[1] When the quantity supplied of the public good is determined through a voting process and individually preferred quantities differ from one another, the collectively chosen quantity will be sensitive to the voting rule, and whether or not efficiency is attained will then also be dependent on the choice of a voting rule.

To ensure neutrality in the supply of the public good with respect to the voting rule, it is necessary that all the individually preferred quantities of the public good be identical—that is to say, $z_1 = z_2 = \ldots = z_r = \bar{z}$.[2] This can be accomplished only by an appropriate distribution of tax prices among the voters. To do this, first set each of the demand functions in (1) equal to $\bar{z}$, and then solve for the corresponding individual tax prices as functions of y and $\bar{z}$:

$$t_i = g(y_i; \bar{z}) \qquad i = 1, \ldots, r \tag{2}$$

Differentiation of (2) yields

$$dt_i = g_y dy_i \qquad i = 1, \ldots, r \tag{3}$$

which shows that the tax prices should vary with y. It is clear then that y is to be the tax base. To obtain a complete characterization of the neutral tax structure, equation (3) is converted into elasticity form:

$$n_{ty} = g_y y / t \tag{4}$$

where n denotes the elasticity of the first subscripted variable with respect to the second and constancy of the elasticities is assumed. Dividing both y and t by $\bar{z}$ on the right-hand side and noting that $g_y = -f_y/f_t$ gives rise to

$$n_{ty} = -n_{zy}/n_{zt} \tag{5}$$

This says that, for the collectively chosen quantity of the public good to be invariant to the change in the voting rule, "The income elasticity of the tax-price schedule must be equal to, but opposite in sign, the income elasticity of demand for the public good divided by the relative price elasticity of demand for the public good" (Buchanan, 1967, p. 153).

Provided the public budget is balanced so that the summed individual tax prices equal the marginal resource cost of the public good, and further provided that a tax imposed on the variable y does not generate an excess burden, a tax structure satisfying (5) is "fully neutral" in Buchanan's terminology:

> In full neutrality the marginal tax-price confronted by each individual is equal to the marginal evaluation that he places on the public good, and the summed tax-prices equal marginal cost. Given these conditions, no individual desires to modify the quantity of the public good that is supplied; all individuals prefer the same quantity. Hence, the political decision rule is immaterial. [1964, p. 228]

This will be recognized as the equivalent of the Lindahl equilibrium in the context of a voting model, with emphasis on neutrality of both the tax and the voting rule.

The requirement that the tax on y not generate an excess burden is unlikely to be met by any known tax system and certainly not by the income tax on which Buchanan's attention was focused. Thus, the direct resource cost of a public good understates the true social cost in that it fails to take account of the inefficiency which the tax imposes on private sector choices. This excess burden is, however, a very real cost of public goods provision to the individual voter-taxpayers. When distortionary taxation is employed, then, individually preferred quantities of the public good will depend upon the sum of tax price and the associated incremental excess burden. This excess burden is often referred to as the welfare cost of taxation, and the marginal welfare cost imposed on voter i will be denoted by w_i.

---

[1] Strictly speaking, only marginal tax price enters the demand function in (1). However, if the public good can be provided at constant unit cost, then it is possible to charge each individual a tax price which is uniform over quantity while still acquiring just enough revenues to cover the full costs of production. In this case, of course, marginal and average tax-prices are equal, and this assumption is adopted here.

[2] It must be noted that, despite appearances, these relationships are not definitionally true since the z_i are individual quantities demanded, not the quantities actually consumed.

Define $m_i = t_i + w_i$ as the marginal cost of the public good to the ith individual. The demand function in (1) now becomes

$$z_i = f(m_i, y_i) \qquad i = 1, \ldots, r \tag{6}$$

Equating all i demand functions and solving for m_i yields

$$m_i = g(y_i; \bar{z}) \qquad i = 1, \ldots, r \tag{7}$$

From (7) can be obtained a formula for the distribution among individuals of the marginal social cost of the public good,

$$n_{my} = -n_{zy}/n_{zm} \tag{8}$$

such that the choice of a voting rule is neutral in terms of its impact on the supply of the public good.

Significant difficulties are encountered in any attempt to measure the aggregate welfare cost imposed on a community by a particular system of taxation, and use of formula (8) requires not just measurement of the aggregate, but a determination of its distribution among individual taxpayers. Thus, it is of some interest to know when Buchanan's original formula in (5) remains valid despite the presence of an excess burden. That is to say, when is formula (8) equivalent to (5)? To see this, expand the left-hand side of (8) and equate with (5) to yield

$$n_{ty}\frac{t}{m} + n_{wy}\frac{w}{m} = n_{ty} \tag{9}$$

which can be solved for $n_{ty} = n_{wy}$. Thus, only if the marginal welfare cost of taxation, which is not a control variable, is distributed among the voter-taxpayers in exactly the same manner as the marginal tax prices will Buchanan's formula continue to be a valid criterion for structuring the tax system to achieve neutrality with respect to the voting rule.

II. Estimating the Efficient Quantity of a Public Good

Under the conditions of public budget balance and no excess burden as specified above, the tax system as described by Buchanan's formula in (5) will result in collective choice of the Lindahl equilibrium quantity of the public good. Moreover, in the absence of an excess burden, the Lindahl equilibrium is known to be efficient as judged by the Pareto criterion. For these reasons it has been suggested by Robin Barlow, Theodore Bergstrom and Robert Goodman, Anthony Barkume, and Howard Pack and Janet Pack that the neutrality formula can be used with estimates of the relevant elasticities of demand for a public good to determine whether or not the quantity of the public good being supplied is efficient. If not, the formula will indicate in what direction and to what extent the quantity actually provided deviates from the optimal quantity.[3]

Suppose, for example, that collective decisions are made by simple majority rule so that the result on any particular decision is the outcome desired by the voter with the median individually preferred outcome. Now, substituting estimates of the demand elasticities for the public good with respect to y and t into (5) yields the corresponding elasticity of the efficient tax-price schedule. If the median of the tax prices prevailing in the community equals the median of the efficient tax prices, then the quantity of the public good supplied by the community is efficient. On the other hand, if the median of the actual tax prices is greater (less) than the median of the efficient schedule, the quantity supplied is less (greater) than the optimum. Moreover, the extent of the divergence can be measured as

---

[3] See also the comments on Barlow by Yoram Barzel and, in particular, by Bergstrom.

$$z^* - \bar{z} = n_{zt}(t_k^* - t_k)\frac{\bar{z}}{t_k} \tag{10}$$

where z^* is the efficient quantity, t^* is the tax price calculated from the efficient schedule, and k indicates the median value.[4]

When the tax system imposes an excess burden, the social cost of the public good exceeds the resource cost so that, as Atkinson and Stern show, the second best optimum will in general require that a smaller quantity of the public good be provided than in the undistorted optimum. If this second best quantity is denoted z', this says that $z' < z^*$. If, now, the proposed test using (5) is applied, there are five possible outcomes, each one leading to an incorrect conclusion. First, suppose that $z' < z^* < \bar{z}$. According to the suggested criterion, the quantity of the public good being supplied to the community exceeds the optimum. This is in fact correct. However, attempting to measure the extent of the divergence using z^* as the optimum would clearly understate the amount of true inefficiency. Next, let $\bar{z} < z' < z^*$. In this case one would correctly conclude that the existing supply of the public good fell short of the efficient quantity. Once again, however, the extent of the divergence would be incorrectly measured, being biased toward overstatement in this case.

In the third case, not only is the extent of the inefficiency likely to be measured incorrectly, but the conclusion drawn is qualitatively erroneous as well. In this case, $z' < \bar{z} < z^*$, and the test suggests that the supply of the public good is too small when in fact it is excessive. Furthermore, only if $|z' - \bar{z}| = |z^* - \bar{z}|$ would the amount of inefficiency be correctly determined. The last two cases arise when $\bar{z} = z^*$, leading to the conclusion that the efficient quantity is being supplied, even though in fact it is excessive, and when $z' = \bar{z}$, in which event one concludes that the supply of the public good is suboptimal, though in truth it is just right.

Unfortunately there appears to be no way to distinguish empirically among these cases in a particular application, except by estimating directly the incremental welfare cost imposed on the median voter. This would permit (8) to be used instead of (5), thereby yielding results which are both qualitatively and quantitatively correct, at least within the restricted arena of the existing tax system. Since different tax systems impose different degrees of excess burden on private sector choices, however, there might exist some other system with similar distribution of the tax burden, but lower aggregate excess burden. Such a tax system would of course be Pareto preferred to the existing system, but the proposed test even using (8) provides no indication of this.

III. Quantitative Dimensions of the Problem

To assess the quantitative dimensions of the problem on which the preceding analysis is focused requires data on the marginal welfare cost of the relevant tax system together with estimates of the price elasticity of demand for the public good. While an extensive investigation of this issue cannot be undertaken here, existing data do permit rough calculations of the potential magnitude of the error and its associated excess burden. Edgar Browning (1976) has estimated the marginal welfare cost of that portion of the *U.S.* individual income tax which falls on labor income under three different assumptions about the particular form which an increase in marginal tax rates might take. As a percentage of the marginal tax revenues required, he found the excess burden to range between 9 and 16 percent.

---

[4] A further assumption is required in order to use (10) as a measure of the extent of divergence from the optimum, specifically, that the median taxpayer has the median individually preferred quantity. This assumption is also necessary in order for the demand estimating procedures used in the cited studies to be valid.

To utilize these estimates, the following assumptions are now made:

1. A unit of public expenditure is that quantity of a public good whose cost (average and marginal) is $1.

2. The existing level of public expenditure is the second best quantity, z'.

3. All incremental revenues required to finance extensions in the supply of public goods would be acquired by increases in the tax rates on labor income only.

With these assumptions, the deviation of the second best quantity from the optimal quantity derived from Buchanan's neutrality formula $z'-z^*$ can be found as the product of marginal welfare cost and the elasticity of demand for public expenditures. If $n_{zt}=-1$, as Browning (1977) assumes, then the potential error in applying Buchanan's formula would itself range between 9 and 16 percent of the existing level of public expenditure. Based on 1974 income tax revenues of approximately $287 billion, this implies that z^* exceeds z' by somewhere between $26 billion and $46 billion. Increasing the level of public expenditure by this amount in accordance with the proposed test would, therefore, generate an excess burden of $1–$3.5 billion. This is over and above the additional excess burden arising from the increase in tax rates.

The magnitude of these distortions is quite sensitive to the value of the elasticity of demand for public expenditures. Bergstrom and Goodman estimate the elasticity of demand for general expenditures of municipal governments to be $-.23$. Using this value as a proxy for the elasticity of demand for federal government expenditures, the range of potential error in applying Buchanan's formula is reduced to 2–3.5 percent of the existing level of expenditure, or between $6 billion and $10 billion. This translates into a potential excess burden of $270 million to $800 million if federal expenditures were increased in accordance with the findings of the proposed test. Of course, if the existing level of public expenditure is not in fact the second best quantity, these estimates may significantly over- or understate the potential error and its associated excess burden. Aggregate data of the kind which currently exists do not permit even crude approximations to be made in this case.

IV. Conclusion

Because of the disincentives to truthful preference revelation, provision of a public good in the private market will almost surely lead to a suboptimal quantity being supplied, with the extent of the inefficiency in all likelihood increasing with the size of the group to be served. Therefore, in the absence of an excess burden arising from taxation, there is a strong presumption in favor of collective provision of the public good. By assigning tax shares to individuals at the constitutional decision stage prior to any consideration of the quantity of the public good to be supplied and by basing these tax shares on some readily observable economic variable which is highly correlated with individual demands for the public good, an indirect link between tax shares and individual demands can be established without providing an incentive for misrepresentation of preferences. The closer the correlation between individual preferences and the chosen tax base, the greater the likelihood that the collectively chosen quantity of the public good will approximate the efficient quantity. With perfect correlation, a tax system structured according to the formula in (5) would result in attainment of an efficient allocation of resources to the public sector through the voting process.

When the excess burden of taxation is taken into account, it has been shown here that the formula in (5) is no longer applicable, but instead the neutral tax structure is characterized by the modified rule in (8). Now, however, the prima facie case for collective provision of a public good vanishes. While private provision leads to inefficiency in the supply of the public good due to the disincentives to truthful preference revelation, collective provision also leads to inefficiency due to the distortionary effect of the tax system on private sector choices, even if the second best tax structure defined by (8) is adopted. Thus, one reaches the conclusion once again that, in seeking methods for improving efficiency in public goods supply, consideration must be given to private market provision as well as to alternative tax systems and mechanisms for inducing truthful revelation of preferences.

REFERENCES

A. B. Atkinson and N. H. Stern, "Pigou, Taxation and Public Goods," *Rev. Econ. Stud.*, Jan. 1974, *41*, 119–28.

A. J. Barkume, "Tax-Prices and Voting Behavior: The Case of Local Educational Financing," *Econ. Inquiry*, Oct. 1977, *15*, 574–86.

R. Barlow, "Efficiency Aspects of Local School Finance," *J. Polit. Econ.*, Sept./Oct. 1970, *78*, 1028–40.

Y. Barzel, "Private Schools and Public School Finance," *J. Polit. Econ.*, Jan./Feb. 1973, *81*, 1974–86.

T. C. Bergstrom, "A Note on Efficient Taxation," *J. Polit. Econ.*, Jan./Feb. 1973, *81*, 187–91.

______ **and R. P. Goodman,** "Private Demands for Public Goods," *Amer. Econ. Rev.*, June 1973, *63*, 280–96.

E. K. Browning, "A Hidden Welfare Cost of Taxation," *Nat. Tax J.*, Mar. 1977, *30*, 89–90.

______, "The Marginal Cost of Public Funds," *J. Polit. Econ.*, Apr. 1976, *84*, 283–98.

James M. Buchanan, "Fiscal Institutions and Efficiency in Collective Outlay," *Amer. Econ. Rev. Proc.*, May 1964, *54*, 227–35.

______, *Public Finance in Democratic Process*, Chapel Hill 1967.

D. B. Johnson and M. V. Pauly, "Excess Burden and the Voluntary Theory of Public Finance," *Economica*, Aug. 1969, *36*, 269–76.

H. Pack and J. R. Pack, "Metropolitan Fragmentation and Local Public Expenditures," *Nat. Tax J.*, Dec. 1978, *31*, 349–62.

P. A. Samuelson, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 357–49.

J. E. Stiglitz and P. Dasgupta, "Differential Taxation, Public Goods and Economic Efficiency," *Rev. Econ. Stud.*, Apr. 1971, *38*, 151–74.

Optimal Public Investment and Dispersion Policy in a System of Open Cities

By ELHANAN HELPMAN AND DAVID PINES*

Cities differ from each other; some differences arise from natural endowments while others from man-made amenities. The quality of a city can be improved by public investment, but at a point of time its quality is historically determined. Individuals tend to choose their residential location in pursuit of utility maximization, and the quality of a city is an important consideration in this locational choice.

Now consider a planner who cannot design the cities *de novo* and cannot construct new cities. He is faced with a given system of cities which he can improve, and he can impose taxes in order to finance public investment in quality improvement. These taxes cannot, however, discriminate among individuals on a personal basis, but they can be different in different locations. This is a rather realistic situation in which the planner is faced with the following problems, which are addressed below (all these problems have been faced, for example, by Israeli planners):

a) How much should be invested in quality improvement, and how should the overall level of investment be distributed among the cities? In particular, should investment take place in (initially) low- or high-quality cities? Should all the cities be brought to the same quality level, or should the terminal quality levels differ among cities?

b) How should public investment projects be financed?

c) How should excessive concentration in high- (or perhaps, low) quality cities be prevented?

d) How should people be encouraged to emigrate to cities located in areas which the planner desires to develop due to noneconomic objectives?

Regarding the first question, which is the main novel issue addressed in this paper, we cannot suggest general rules. The precise answer depends on many economic variables like preferences, initial qualities of cities, transportation functions, etc. However, we argue that there are circumstances in which it is optimal to invest only in the highest-quality city and there are circumstances in which it is optimal to invest only in the lowest-quality cities. This argument is substantiated with an example which is fully developed in our discussion paper.

Some of the problems discussed in the present paper are related to problems discussed in the paper by Frank Flatters, Vernon Henderson, and Peter Mieszkowski. They have characterized the optimal allocation of resources, including local public goods, in a system of communities with population mobility, and so do we (the quality of a city is a local public good). But their focus was on the case in which there are no restrictions (except for resource constraints) on the provision of local public goods, while we consider the case in which communities have different initial amenities which can only be increased. This enables us to discuss optimal public investment in relation to the initial distribution of qualities among cities —an important problem that they did not address. In addition, in contrast to their model, we employ a disaggregated framework in terms of the intraurban structure, which enables us to provide valuable insight into the problems at hand. However, when it comes to the same issues, our findings confirm theirs. Thus, a development policy (the supply of public goods in their case) designed to improve the quality levels cannot be delegated to local governments, but should be centrally coordinated, and an

---

*Department of economics, Tel-Aviv University. This work was completed when Helpman was on leave at the University of Rochester. We wish to thank Richard Arnott for comments.

efficient financing system of public investment consists of equal lump sum taxes on each household irrespective of its location, as long as there are no noneconomic objectives in the dispersion of the population. This tax system is also required to guarantee the optimal distribution of the population by preventing excessive concentration in high-quality (or, perhaps, low-quality) cities.

We present our model in Section I, in which we derive also the first-order conditions for an optimal allocation. In Section II, we describe properties of an optimal allocation, including optimal public investment strategies. Section III concludes the paper with a discussion of optimal policies for the attainment of a desired population dispersion policy that arises from noneconomic objectives.

I. The Model

Consider a system of m monocentric cities which differ from each other by their amenity levels. These may represent climatic properties, scenic features, or other natural endowments. But they can also represent historic man-made amenities such as the layout of roads, parks, or impressive statues. It is assumed that the amenity level of a given city can be improved by public investment, but that new cities cannot be constructed. Accordingly, the terminal quality level of city i, Q_i, depends on the initial attributes of the city, A_i, and on the level of public investment in improving amenities in city i, P_i. Formally:

$$Q_i = q_i(A_i, P_i) \qquad i = 1,2,\dots,m$$

where A_i = vector of initial attributes of city i, and P_i = public investment in city i.

We assume that $q_{iA}, q_{iP} > 0$, where $q_{iX} = \partial q_i / \partial X_i$, $X = A, P$. Let

$$(1) \qquad P_i = p_i(Q_i, A_i) \qquad i = 1,2,\dots,m$$

be the inverse function of $q_i(\cdot)$, then $p_{iQ} > 0$ where $p_{iQ} = \partial p_i / \partial Q_i$.

The population of the whole system of cities is assumed to be homogeneous in terms of both preferences and initial holdings. Homogeneity in terms of initial holdings implies that either (a) landlords are absentees, or (b) each household owns equal shares of land in every location, and is therefore entitled to equal shares in the rent.

We prefer the second assumption, since some inconsistency is inherent in the first one. If the absentee landlords are accounted for in optimality considerations, then a priori, households cannot be said to be homogeneous. The reason is that some of them own initial endowments which others do not. If, on the other hand, the absentee landlords are not accounted for in optimality considerations, then it is not clear why their land is not confiscated by appropriate land taxes which are redistributed as lump sum transfers. This, of course, implies that the second assumption is satisfied. Thus, we follow the second assumption rather than the first.

The utility of each household is assumed to be a quasi-concave function of housing $H(x)$, the consumption of all other commodities $Z(x)$ (referred to hereafter as a composite commodity), and "locational quality" Q_i.

$$(2) \qquad U_i(x) = u[H_i(x), Z_i(x), Q_i] \qquad i = 1,2,\dots,m$$

where $U_i(x)$ = utility level of a household residing x miles from the center of city i. We assume $u_H, u_Z, u_Q > 0$, where $u_X = \partial u / \partial X_i, X = H, Z, Q$.

Assuming that the composite commodity is the numeraire, the compensated demands for housing and the composite commodity are

$$(3) \qquad H_i(x) = h[Q_i, R_i(x), U_i(x)] \qquad i = 1,2,\dots,m$$

$$(4) \qquad Z_i(x) = z[Q_i, R_i(x), U_i(x)] \qquad i = 1,2,\dots,m$$

where $R_i(x)$ = rent per unit of housing x miles from the center of city i.

Without loss of generality, it is assumed that housing is produced by land only. Thus, if $\theta_i(x)>0$ represents the amount of land available x miles from the center of city i, we have

$$n_i(x)H_i(x)\leqslant\theta_i(x)\qquad i=1,2,\ldots,m \tag{5}$$

where $n_i(x)$ is the number of households residing x miles from the center of city i.

Let the total amount of nonland resource be $\bar{Z}$. These resources are used for direct consumption, transportation, rent to agricultural landlords (exogeneously determined), and public investment. Accordingly,

$$\sum_{i=1}^{m}\int_0^{S_i}[\{Z_i(x)+t_i(x)\}\cdot n_i(x)+\theta_i(x)R_i^A]dx+\sum_{i=1}^{m}p_i(Q_i,A_i)\leqslant\bar{Z} \tag{6}$$

where S_i = the distance from the center of city i to its most distant point, $t_i(x)$ = transportation costs (an increasing function of x), and R_i^A = alternative cost of land put into urban use (agricultural rent).

The population to be distributed among the cities is $\bar{N}$, which implies

$$\bar{N}=\sum_{i=1}^{m}\int_0^{S_i}n_i(x)dx \tag{7}$$

The policy variables comprise prices, nondiscriminating taxes, and the allocation of public investment among cities. The choice of prices as policy variables is a convenience which is justified by the fact that every desired price structure can be achieved by taxes which do not discriminate among individuals. These prices (taxes) have of course to be chosen from a set of equilibrium price configurations. A similar procedure is adopted in much of the optimal taxation literature.

Equilibrium under the above specified assumptions implies equal treatment of equals. Hence, we have

$$U_i(x)=U\qquad i=1,2,\ldots,m \tag{8}$$

Since we do not permit negative public investment, we must have

$$Q_i\geqslant q_i(A_i,0)\qquad i=1,2,\ldots,m \tag{9}$$

The optimal allocation problem is now defined as follows: Choose U,Q_i,S_i, and $n_i(x),R_i(x)$ for $x\in[0,S_i]$, $i=1,\ldots,m$ to maximize U subject to equations (3)–(9) and $S_i,n_i(x),R_i(x)\geqslant 0$, $i=1,2,\ldots,m$.

At this stage we want to characterize the optimal policies and to compare the structure of cities which are *occupied* in an optimal allocation; that is, cities in which it is optimal to have a strictly positive population size. For this purpose let M (a subset of $\{1,2,\ldots,m\}$) be the set of cities which are occupied at the optimal allocation. Then, we can characterize the optimal allocation by characterizing the solution to an optimization problem which is similar to the one stated above, except that the index i runs now over the elements of M instead of $\{1,2,\ldots,m\}$. In the new problem the constraints $S_i,R_i(x)\geqslant 0$ for $x\in[0,S_i]$, $i\in M$, cannot be binding and the constraint $n_i(x)\geqslant 0$ for $x\in[0,S_i],i\in M$, cannot be binding at points x at which $\theta_i(x)>0$. By considering the modified problem instead of the original one, we avoid the discussion of how to optimally choose the subset of cities with a positive population size, a subject with which we are not concerned in this paper.

The following is the Lagrangian that corresponds to the modified problem:

$$\begin{aligned}L=U&+\lambda\Big\{\bar{Z}-\sum_{i\in M}\int_0^{S_i}[\{z[Q_i,R_i(x),U]+t_i(x)\}n_i(x)+\theta_i(x)R_i^A]dx-\sum_{i\in M}p_i(Q_i,A_i)\Big\}\\&+\mu\Big[\sum_{i\in M}\int_0^{S_i}n_i(x)dx-\bar{N}\Big]\\&+\sum_{i\in M}\int_0^{S_i}\rho_i(x)\{\theta_i(x)-h[Q_i,R_i(x),U]n_i(x)\}dx\\&+\sum_{i\in M}\gamma_i[Q_i-q_i(A_i,0)]\end{aligned} \tag{10}$$

where λ, μ, $\rho_i(x)$, and γ_i are the Lagrangian multipliers which correspond to (6), (7), (5), and (9), respectively, augmented by constraints (3), (4), and (8).

The necessary first-order conditions for a maximum are

$$(11)\quad \partial L/\partial U = 1-\lambda \sum_{i\in M}\int_0^{S_i} z_U(i,x)n_i(x)dx - \sum_{i\in M}\int_0^{S_i}\rho_i(x)h_U(i,x)n_i(x)dx=0$$

$$(12)\quad \partial L/\partial Q_i = -\lambda\int_0^{S_i} z_Q(i,x)n_i(x)dx-\lambda p_{iQ}(Q_i,A_i) - \int_0^{S_i}\rho_i(x)h_Q(i,x)n_i(x)dx+\gamma_i=0 \quad \text{for } i\in M$$

$$(13)\quad \partial L/\partial S_i = -\lambda\{[z(i,S_i)+t_i(S_i)]n_i(S_i) + \theta_i(S_i)R_i^A\}+\mu n_i(S_i)=0 \quad \text{for } i\in M$$

$$(14)\quad -\lambda[z(i,x)+t_i(x)]+\mu-\rho_i(x)h(i,x)=0 \quad \text{for } x\in[0,S_i],\quad i\in M$$

$$(15)\quad -\lambda z_R(i,x)-\rho_i(x)h_R(i,x)=0 \quad \text{for } x\in[0,S_i],\quad i\in M$$

where (14) has to be satisfied for optimality of $n_i(x)$; (15) has to be satisfied for optimality of $R_i(x)$; and $e(i,x)\equiv e[Q_i,R_i(x),U]$ for $e=z,h$.

Condition (15) and the properties of compensated demand functions imply

$$(16)\quad R_i(x)=\rho_i(x)/\lambda \quad \text{for } x\in[0,S_i],\ i\in M$$

Equations (14) and (16) yield

$$(17)\quad z(i,x)+R_i(x)h(i,x)+t_i(x)=\mu/\lambda \quad \text{for } x\in[0,S_i],\ i\in M$$

Hence, μ/λ is optimal disposable income.

II. Properties of an Optimal Allocation

We start by describing some general characteristics of an optimal allocation. The first proposition describes the optimal distribution of disposable income. From equation (17), we have:

PROPOSITION 1: *In an optimal allocation all individuals have the same disposable income, irrespective of their location; that is, disposable income varies neither within nor among cities.*

Proposition 1 implies that public investment should be financed by lump sum taxes, or other instruments which are equivalent to lump sum taxes. Moreover, since the level of investment in city i is not related to the tax bill of city i, this implies that optimality may require intercity transfers and central coordination.

In order to derive more results, we need more assumptions and some further rewriting of the first-order conditions. The additional results are based on different degrees of similarity among the cities.

Substituting (16) and constraint (5) with strict equality into (14) (constraint (5) has to be satisfied with strict equality because $\rho_i(x)>0$ is necessary for the satisfaction of the first-order conditions) and comparing the result for $x=S_i$ with condition (13) yields

$$(18)\quad R_i(S_i)=R_i^A \quad \text{for } i\in M$$

We can also rewrite (17) in the form

$$(19)\quad y[Q_i,R_i(x),U]=\frac{\mu}{\lambda}-t_i(x) \quad \text{for } x\in[0,S_i],\ i\in M$$

where $y(\cdot)$ is the minimum expenditure function, defined for expenditure on the composite good and housing; $y_Q<0$, $y_{QQ}>0$, $y_R=h$. Using (18), we also have

$$(20)\quad y[Q_i,R_i^A,U]=\frac{\mu}{\lambda}-t_i(S_i) \quad \text{for } i\in M$$

We make now the assumption that the alternative value of land is the same in all occupied cities and that they have the same transportation functions in the relevant range.

ASSUMPTION 1: $R_i^A = R^A$ for all $i \in M$, and $t_i(x) = t(x)$ for all $x \in [0, \max_{i \in M} S_i]$, $i \in M$.

This assumption enables us to derive the following proposition:

PROPOSITION 2: *Under Assumption* 1, *in an optimal allocation*: (a) *Cities with higher-terminal quality are at least as large as cities with lower-terminal quality in the sense that the longest distance (from the center) of inhabited areas is at least as great in cities with higher-terminal quality as in cities with lower-terminal quality.* (b) *The rent at any given distance from the center is at least as high in cities with higher-terminal quality than in cities with lower-terminal quality.*

PROOF:

Given $R_i^A = R^A$ and $t_i(\cdot) = t(\cdot)$, we can solve from (20): $S_i = \hat{s}(Q_i, R^A, U, \mu/\lambda)$. Since $y(\cdot)$ decreases with Q_i and t increases with S_i, $\hat{s}(\cdot)$ increases with Q_i. This proves (a). Using (19), we can solve: $R_i(x) = \hat{r}[Q_i, U, \mu/\lambda - t(x)]$. Since $y(\cdot)$ decreases with Q_i and increases with $R_i(x)$, $\hat{r}(\cdot)$ increases with Q_i, which proves (b).

The above proof implies that under Assumption 1, there exists functions $s(Q_i)$ and $r(Q_i, x)$ such that $s(\cdot)$ and $r(\cdot)$ increase with Q_i and $r(\cdot)$ declines with x, and at the optimal solution:

$$S_i = s(Q_i) \equiv \hat{s}\left(Q_i, R^A, U, \frac{\mu}{\lambda}\right) \tag{21}$$

$$R_i(x) = r(Q_i, x) \equiv \hat{r}\left[Q_i, U, \frac{\mu}{\lambda} - t(x)\right] \tag{22}$$

where U and μ/λ are at their optimal values. We add now two further assumptions.

ASSUMPTION 2: $\theta(x) = \theta_i(x)$ for all $i \in M, x \in [0, \max_{i \in M} S_i]$.

ASSUMPTION 3: Housing and locational quality are net substitutes; i.e., $h_Q < 0$.[1]

PROPOSITION 3: *Under Assumptions* 1–3, *in an optimal allocation*: (a) *For every distance from the center, density is at least as high in cities with higher-terminal quality than in cities with lower-terminal quality.* (b) *Total population is larger in cities with higher-terminal quality than in cities with lower-terminal quality.*

PROOF:

Under Assumptions 1–3, we have

$$h[Q_i, R_i(x), U]\, n_i(x) = \theta(x) \qquad i \in M$$

$$R_i(x) = r(Q_i, x) \qquad i \in M$$

Since $h(\cdot)$ decreases with Q_i and $R_i(x)$ (h_R is the pure substitution effect), and $r(\cdot)$ increases with Q_i, then for fixed x, U: $\hat{h}(Q_i, U, x) \equiv h[Q_i, r(Q_i, x), U]$ declines with Q_i. Hence, $n_i(x) = \hat{n}(Q_i, U, x) \equiv \theta(x)/\hat{h}(Q_i, U, x)$ increases with Q_i, and so does $\hat{n}(Q_i, U, x)/\theta(x)$. The last expression represents density. This proves part (a).

To prove part (b), consider a city's population

$$N(Q_i) \equiv \int_0^{s(Q_i)} \hat{n}(Q_i, U, x)\, dx \qquad i \in M$$

Since $\hat{n}(\cdot)$ and $s(\cdot)$ increase with Q_i, $N(\cdot)$ increases with Q_i. This proves part (b).

We may, in fact, represent population at a given distance from the center as a function of terminal quality and distance.

$$n_i(x) = n(Q_i, x) \equiv \hat{n}(Q_i, U, x) \tag{23}$$

where $\hat{n}(\cdot)$ is defined in the proof of Proposition 3 and $n(\cdot)$ increases with Q_i. Now, using (16), (12) can be rewritten as

$$B_i(Q_i) \equiv -\int_0^{S_i} y_Q[Q_i, R_i(x), U]\, n_i(x)\, dx = p_{iQ}(Q_i, A_i) - \gamma_i \tag{24}$$

where S_i, $R_i(x)$, U, and $n_i(x)$ are optimal.

The function $B_i(Q_i)$ represents the marginal social benefit of increasing city i's quality; $p_{iQ}(Q_i, A_i)$ is the social marginal

---

[1] $h_Q < 0$ if and only if housing and quality are net substitutes in the Hicks-Slutzky sense.

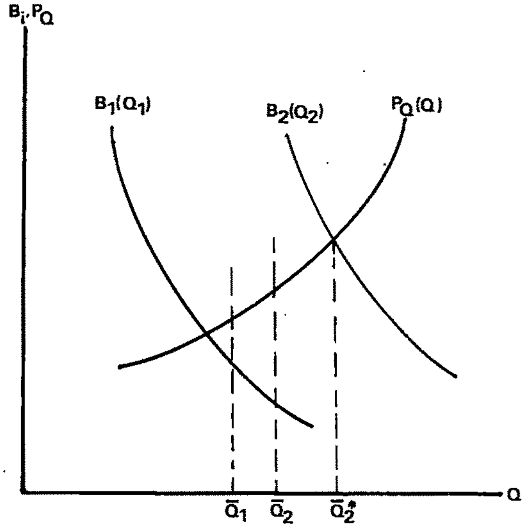


FIGURE 1

cost. If $\gamma_i > 0$, then $P_i = 0$; i.e., if marginal costs exceed marginal benefits at $P_i = 0$, no public investment should take place in city i. The condition $P_i > 0$ implies $\gamma_i = 0$; i.e., if it is profitable to undertake public investment in city i, it should be carried out to the point where marginal social benefits equal marginal social costs. Condition (24) can also be interpreted as the well-known Samuelson condition for optimal provision of public goods, since y_Q equals the private marginal rate of substitution between Q and Z.

It is easy to show that y is convex in Q_i. Hence, $B_i(\cdot)$ declines in Q_i. Now, assume that $p_{iQ}(\cdot)$ does not depend on A_i, that it increases with Q_i, and that it is the same for all $i \in M$. Figure 1 describes the optimal investment levels for this case. In Figure 1, we have two marginal benefit curves, $B_1(\cdot)$ and $B_2(\cdot)$. They are drawn on the assumption that we use the correct (terminal) values of $S_i, R_i(x), U$. Denoting $\bar{Q}_i = q(0, A_i)$, optimality implies that one should *not* invest in city 1, so that $Q_1 = \bar{Q}_i$ is the optimal value of Q_1, while one should invest in city 2 up to the point where $Q_2 = Q_2^*$.

We have thus characterized optimal public investment in terms of cost-benefit analysis. However, this characterization does not reveal the optimal public investment strategy, in the sense that it does not reveal conditions under which optimal public investment is related to initial qualities of cities. Should we invest in initially low, medium, or high-quality cities? In Israel, for example, the tendency was to invest in low-quality cities.

Using our general model, we have not been able to derive conditions under which it is optimal to invest in initially low-quality cities or initially high-quality cities. It can be argued, however, and the argument can be supported by examples, that both cases may occur. The intuitive argument is as follows. Suppose that our cities are identical in all respects, except for their initial quality levels. Then clearly, if there are no transportation costs, it is optimal to concentrate the entire population in the highest-quality city and to improve its quality level up to the point at which the Samuelson condition is satisfied (remember that the supply of land is infinitely elastic.) This shows that it may be optimal to invest only in the city with the highest initial quality level. However, in the presence of substantial transportation costs, this need not be the optimal strategy, because in this case the dispersion of the population among more cities reduces average transportation costs. Transportation imposes an indirect cost on housing services. Therefore, the more important is housing in consumer preferences and the less important is quality, the more even should be the distribution of the population among cities with different quality levels. But as low-quality cities become more populated, the marginal benefit of investing in these cities rises. Under these circumstances it may be optimal to invest in low-quality cities.

The preceding argument can be substantiated by means of an example (developed in more detail in our discussion paper), where cities are linear and they differ only in their initial quality levels. The transportation functions are linear and preferences are Cobb-Douglas. The alternative cost of land is zero, and preferences are normalized so that the sum of the coefficients on the composite good and housing equals one. In this specific case the investment strategy depends only on the elasticities of the utility function with respect to housing and city quality. If the elasticity of the utility function with respect to quality exceeds the elasticity with respect to housing, *the optimal strategy is to invest only in one city—the city with the initially highest-quality level.* If, on the other hand, the elasticity of the utility function with respect to housing exceeds its elasticity with respect to quality, *the optimal strategy is to invest in the lowest-quality cities and to bring them to a common terminal quality level which does not exceed the quality levels of the remaining cities in the system.* In an optimal allocation all cities with initially positive quality levels are populated.

III. Optimal Dispersion Policies

Population dispersion policies are an important issue in many countries. Countries may have preferences over alternative distributions of their population among settlements—preferences which may stem from political, social, or security considerations; Israel is a case in point. In order to achieve these goals, policymakers can employ direct measures (such as preferential income taxation, housing subsidies, etc.) or indirect measures (such as more intensive public investment in the preferred cities); all these measures are employed in Israel. What is the optimal policy mix? This question can be answered by slightly modifying the model of Section I.

Let the policymaker's objective function be $W(U,N_1,N_2,\dots,N_m)$ where N_i is the population of city i. This is a very general objective function which increases with U, and which may depend in a complex way on the distribution of the population among cities.

Now, rewrite constraint (7) as

$$N_i=\int_0^{S_i} n_i(x)dx \qquad i=1,2,\dots,m \tag{7'}$$

$$\bar{N}=\sum_{i=1}^{m} N_i \tag{7''}$$

Our new optimization problem is: Choose U,Q_i,S_i,N_i, and $n_i(x),R_i(x)$ for $x\in[0,S_i], i=1,2,\dots,m$, to maximize $W(U,N_1,N_2,\dots,N_m)$ subject to (3), (4), (5), (6), (7′), (7″), (8), (9), and $S_i,N_i,n_i(x),R_i(x)\geqslant 0$.

The corresponding Lagrangian is

$$\begin{aligned} L= {} & W(U,N_1,N_2,\dots,N_m)+ \\ & \lambda\Big\{\bar{Z}-\sum_{i\in M}\int_0^{S_i}\big[\{z[Q_i,R_i(x),U]+ \\ & t_i(x)\}n_i(x)+\theta_i(x)R_1^A\big]dx+\sum_{i\in M}p_i(Q_i,A_i)\Big\} \\ & +\mu\Big[\sum_{i\in M}N_i-\bar{N}\Big]+\sum_{i\in M}\mu_i\Big[\int_0^{S_i}n_i(x)dx-N_i\Big] \\ & +\sum_{i\in M}\int_0^{S_i}\rho_i(x)\{\theta_i(x)-h[Q_i,R_i(x),U] \\ & \cdot n_i(x)\}dx+\sum_{i\in M}\gamma_i[Q_i-q_i(A_i,0)] \end{aligned} \tag{10'}$$

where μ and μ_i are the Lagrangian multipliers of (7′) and (7″), respectively.

It can be shown that the only qualitative difference between the solution of the previous problem and the solution of the present problem is that now disposable income may differ among cities (but not within cities), and μ_i/λ is the optimal disposable income for residences of city i. From the optimal choice of N_i, we have

$$\mu_i=\mu+\partial W/\partial N_i \tag{25}$$

Hence, if cities i and j are such that the planner does not care (possible at the margin) about their populations ($\partial W/\partial N_i=\partial W/\partial N_j=0$), their residences should have the same disposable income ($\mu_i=\mu_j=\mu$). However, whenever the social marginal utility of city i's population differs from the social margin utility of city j's population ($\partial W/\partial N_i\neq\partial W/\partial W_j$), income should differ between these cities ($\mu_i-\mu_j=\partial W/\partial N_i-\partial W/\partial N_j$).

It is clear from (25) that the optimal population dispersion policy is a location subsidy to households. The subsidy should be positive in city *i* if the marginal social utility of city *i*'s population is positive, and it should be negative (a tax) if the marginal social utility of city *i*'s population is negative. The policymaker should not interfere in the housing market in order to achieve an optimal distribution of the population. It is also clear that a given dispersion of the population can be brought about by an appropriate investment pattern. But investment policy alone is insufficient for implementing the optimal policy; a scheme of lump sum transfers and income redistribution has to be applied.

REFERENCES

F. Flatters, V. Henderson, and P. Mieszkowski, "Public Goods, Efficiency and Regional Fiscal Equalization," *J. Publ. Econ.*, Aug. 1974, *3*, 99–112.

E. Helpman and D. Pines, "Optimal Public Investment and Dispersion Policy in a System of Open Cities," disc. paper no. 78-8, department of economics, Univ. of Rochester, Sept. 1978.

Gains from Trade under Uncertainty: Further Comment

By HIROYA AKIBA*

In an article appearing in this *Review*, Raveendra Batra and William Russell (hereafter B-R) analyze the effect of uncertainty in the international terms of trade on social welfare within the framework of a small country with two goods. They conclude that the expected social welfare decreases under both first- and second-degree stochastic dominance. Also in a recent issue of this *Review*, Richard Hartman, commenting on B-R's article, contends that (i) B-R do not allow for the change in the optimal consumption of one good, c_1, as the probability distribution of the international terms of trade p undergoes a mean-preserving spread, and (ii) they do not take into account the change in c_1 when comparing the expected utility for different distributions of p. Based on these criticisms, Hartman presents an alternative proof to show a decrease in the expected social welfare when the distribution of p undergoes a mean-preserving spread.

Batra and Russell are interested in the change of the expected social welfare when the distribution of p changes. In their model, c_1 is chosen before the uncertainty in p is resolved. To analyze the effect of uncertainty in p on social welfare, B-R do not allow for the change in the optimal c_1 under different distributions of p on the ground that c_1 is chosen optimally before p is known. However, as pointed out by Hartman, the optimal value of c_1 will be different for different distributions of p. To discuss the change in the relationship between the expected social welfare and the optimal c_1, Hartman uses a diagram similar to my Figure 1. From his diagram Hartman argues that B-R's conclusion of the decrease in the expected social welfare under greater uncertainty is correct even if the optimal c_1 is different for different distribution of p.[1]

In my view, B-R, as well as Hartman, only consider a special case where the expected utility curve under a probability distribution always lies above those under different probability distributions of p. If we accept Hartman's comments, I can think of another case where the expected utility curves under different probability distributions may cross. The purpose of this note is first to examine this case that could occur within the framework of the B-R analysis, but which has not been examined by B-R or Hartman. By failing to examine this case, the B-R and Hartman analyses seem incomplete. Also, by considering this case, I modify the B-R proof to show a decrease in the expected social welfare as the distribution of p undergoes a mean-preserving spread.

First, Hartman's analysis of the B-R conclusion is presented. This is followed by an examination of the other possible situation. Finally, the modification of the proof is presented. The discussion in this note is confined to the B-R framework and their notation is used.

The B-R proof simply provides a stronger sufficient condition for the decrease in the expected social welfare and does not rely on the usual comparative statics approach. In Figure 1, I denote the original equilibrium position by point 1 on the expected utility (EU) curve, I.

Batra-Russell prove that for every c_1, EU decreases when the exogenous changes in distribution of the international terms of

---

*Takushoku University. I acknowledge Vincent Munley and Kambiz Kiani for their comments on an earlier version of this note. I am also indebted to an anonymous referee and Harry Ramcharran for their comments and suggestions. Any remaining errors are my own.

[1] In his alternative proof summarized in his equation (4) on page 927, Hartman treats c_1 as constant as B-R do. This contradicts his criticism of their article.

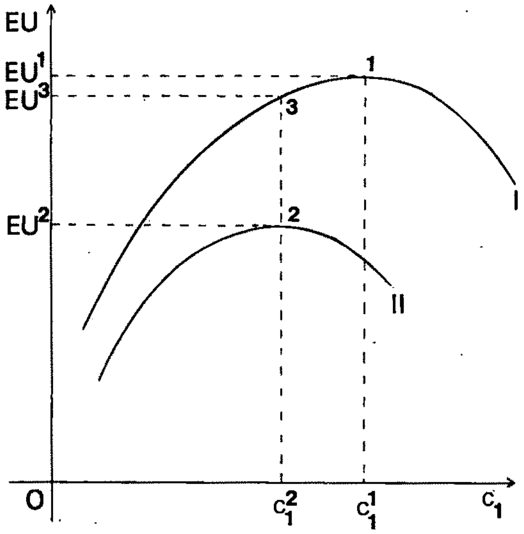


FIGURE 1

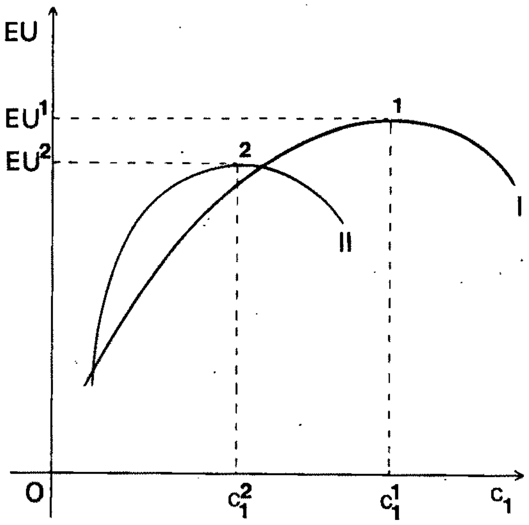


FIGURE 2

trade is introduced. That is, the curve shifts down to *II*. Whatever is the new equilibrium of c_1 on curve *II* (see point 2), there is some point on curve *I* (see point 3) which lies above it. Thus, $EU^3 > EU^2$. But since point 1 was initially the equilibrium point (and chosen over point 3) we have $EU^1 > EU^3 > EU^2$. It is obvious that once a decrease in EU is established for every c_1 we must accept that result for the new equilibrium c_1^2 as well (see B-R, 1976).

However, once we allow for the fact that the optimal c_1 will generally be different for different distributions of the terms of trade, we may have a situation where the expected utility curves may cross (see Figure 2). Then both B-R's and Hartman's proofs fail to show that $EU^1 > EU^2$. This will be shown in what follows. Notation is defined in the Appendix.

As the probability density changes from g_1 to g_2, the change in the expected social welfare is expressed as[2]

$$\Delta E[U] = \int_0^\infty U(c_1, c_2)\{g_2(c_2) - g_1(c_2)\}dc_2$$

which is B-R's (1974) equation (9). Integrating by parts by allowing for the change in c_1 explicitly, we have

$$(1)\quad \Delta E[U] = [U(c_1,c_2)\{G_2(c_2) - G_1(c_2)\}]_0^\infty - \int_0^\infty (U_1 dc_1/dc_2 + U_2)\{G_2(c_2) - G_1(c_2)\}dc_2$$

$$= \int_0^\infty \left(\frac{U_1 - pU_2}{p}\right)\{G_2(c_2) - G_1(c_2)\}dc_2$$

where U_i $(i = 1,2)$ are evaluated under the original distribution function, $G_1(c_2) = \int_0^{c_2} g_1(t)dt$. Then, if we notice that[3]

$$(2)\quad d\left(\frac{U_1 - pU_2}{p}\right) \Big/ dc_2 = \frac{1}{p^2}(-U_{11} + 2pU_{12} - p^2U_{22}) > 0$$

---

[2] I agree with Hartman for taking zero and infinity as the limits of integration.

[3] The second-order condition for the expected utility maximization with respect to c_1 is $d^2E[U]/dc_1^2 = E[U_{11} - 2pU_{12} + p^2U_{22}] < 0$. Batra-Russell's assumption, $U_{ii} < 0$ $(i = 1,2)$, is not enough for this to be satisfied. One sufficient but not necessary condition for this is $U_{12} > 0$. Under these assumptions (2) is positive.

equation (1) may be seen as negative for $G_2(c_2) \geq G_1(c_2)$.[4]

Upon integration of (1) once again, we have

$$(3)\quad \Delta E[U] = \left[\left(\frac{U_1 - pU_2}{p} \right) \times \left\{ \int_0^{c_2} G_2(t)dt - \int_0^{c_2} G_1(t)dt \right\} \right]_0^{\infty} - \int_0^{\infty} (1/p)[U_{11}dc_1/dc_2 + U_{12} - pU_{12}dc_1/dc_2 - pU_{22}] \times \left\{ \int_0^{c_2} G_2(t)dt - \int_0^{c_2} G_1(t)dt \right\} dc_2$$

$$= \int_0^{\infty} \frac{1}{p^2}(U_{11} - 2pU_{12} + p^2U_{22}) \times \left\{ \int_0^{c_2} G_2(t)dt - \int_0^{c_2} G_1(t)dt \right\} dc_2$$

Equation (3) is evaluated under the condition of a mean-preserving spread (see B-R, 1974, p. 1043). Then in view of our assumption (2), we have $\Delta E[U] < 0$ for $\int_0^{c_2} G_2(t)dt \geq \int_0^{c_2} G_1(t)dt$.

In summary, the Batra-Russell conclusion that the increased uncertainty in the terms of trade leads to a decline in the expected utility is shown to be correct by taking an explicit account of the fact that the different probability distributions yield different optimal values of c_1. The formal proof of this fact is implied in a recent contribution by Giora Hanoch in a more general analysis of price uncertainty, although Hartman mentioned it loosely. The alternative proof presented by Hartman does not properly take this into account.

APPENDIX

$c_i(i=1,2)$ = consumption of the ith commodity

P = the international terms of trade (p_1/p_2) $(p_i(i=1,2)$ = price of the ith commodity)

EU = the expected utility

$g_i(i=1,2)$ = the probability density function of c_2

$G_i(i=1,2)$ = the cumulative distribution function of c_2

$U(c_1,c_2)$ = a social utility function

$\Delta E[U]$ = the change in the expected utility

$U_i(i=1,2)$ = the partial derivative of U with respect to c_i

$U_{ij}(i,j=1,2)$ = the second partial derivative of U with respect to c_i and c_j

REFERENCES

R. N. Batra and W. R. Russell, "Gains From Trade Under Uncertainty," *Amer. Econ. Rev.*, Dec. 1974, *64*, 1040–48.

______ **and** ______, "Gains From Trade Under Uncertainty: Reply," *Amer. Econ. Rev.*, Dec. 1976, *66*, 928.

G. Hanoch, "Risk Aversion and Consumer Preferences," *Econometrica*, Mar. 1977, *42*, 413–26.

R. Hartman, "Gains From Trade Under Uncertainty: Comment," *Amer. Econ. Rev.*, Dec. 1976, *66*, 925–27.

---

[4] It should be pointed out, however, that the above argument is true if $E[U]$ is maximized under the original distribution function, $G_1(c_2)$. In this sense, the result obtained here may not be considered as a strict interpretation of the first-degree stochastic dominance.

The Comparative Statics of the Competitive, Increasing-Cost Industry

By JOSEPH P. HUGHES*

It is commonly thought that an increase in the demand for the output of a perfectly competitive, increasing-cost industry stimulates the entry of new firms, and that without such entry the existing firms would attain long-run equilibrium with profits. For example, Edwin Mansfield notes in a well-known textbook that the result of an outward shift in demand is that "...the price of the product increases and firms earn economic profits, thus attracting new entrants. More and more inputs are required by the industry, and in an increasing cost industry, the price of inputs increases with the amount used by the industry" (pp. 247–48). Similarly, C. E. Ferguson and John Gould conclude, "The number of firms and industry output must increase. However, there is no way to predict what will happen to the equilibrium output per firm.... But these items are certain: industry output, the number of firms, and the long-run supply price will increase" (p. 252).

Unfortunately, in failing to consider more carefully the response of the firm's equilibrium output, these analyses are misleading, and in certain cases incorrect: an increase in demand may induce losses and exit as well as profits and entry. Simplifying the analysis by assuming the industry faces a rising supply curve in only one factor market and that this constitutes the source of the industry's increasing costs, this paper proves that when that factor is normal, an increase in its price caused by the expansion of industry output may raise average cost more than the increase in demand raises average revenue. Hence, firms experience losses, some leave the industry, and the higher factor price induces the remaining firms to produce at a larger scale. Thus, losses and exit as well as profits and entry may be the result of an increase in demand. The failure to account adequately for scale adjustments turns out to be the source of the common misconception.

I. Changes in Demand and the Number of Firms

Consider a perfectly competitive firm producing the rate of output q with n inputs employed at the rates $x=(x^1,\ldots,x^n)$ and hired at the prices $w=(w^1,\ldots,w^n)$. It is assumed that the firm's cost function $C(q,w)$ issues U-shaped average cost curves for any given w.[1] These additional definitions and familiar relations will be needed.

(i) Shephard's Duality Theorem[2] demonstrates that the cost-minimizing demand for an input can be determined from the derivative

$$C_j(q,w)=x^j(q,w) \tag{1}$$

where $C_j=\partial C/\partial w^j$.

(ii) The concavity of $C(q,w)$ in w implies that the cost-minimizing demand for an input is inversely related to its own price (i.e., the substitution effect is nonpositive):

$$C_{jj}(q,w)=x_j^j(q,w)\leqq 0 \tag{2}$$

where, again subscripts indicate differentiation.

(iii) Assuming the cost function is twice continuously differentiable, Young's Theorem notes that

$$C_{jq}=C_{qj} \tag{3}$$

---

*Assistant professor of economics, Rutgers College. I wish to thank George Borts and an anonymous referee for their generous assistance.

[1] Giora Hanoch discusses the technical requirements of the production function that are necessary and sufficient for U-shaped average cost curves. See his Proposition 1 and Corollary 1.

[2] For a brief discussion of Shephard's Duality Theorem, see W. Erwin Diewert.

and from (1) it is clear that

$$C_{jq} = x_q^j \tag{4}$$

so that

$$x_q^j = C_{qj} \tag{5}$$

(iv) Where profit is maximized, $C_{qq} > 0$.

(v) Defining the output elasticity of an input along the expansion path as

$$\eta^j = (q/x^j)x_q^j \tag{6}$$

the jth input is said to be superior when $\eta^j > 1$, normal when $0 < \eta^j < 1$, and inferior when $\eta^j < 0$. Thus, from (5) it can be concluded that marginal cost varies inversely with the price of the jth input only when it is inferior.

(vi) Assume the industry is composed of N identical firms and faces perfectly elastic supplies of all inputs but the ith. Let $g(w^i)$ represent the supply of the ith input to the industry. Although the industry experiences a rising supply curve, $g_i(w^i) > 0$, the individual firm's supply is perfectly elastic at w^i.

The long-run equilibrium of the firm and industry are summarized in the conditions

$$p - C_q(q, w) = 0 \tag{7}$$

$$Nx^i(q, w) - g(w^i) = 0 \tag{8}$$

$$pq - C(q, w) = 0 \tag{9}$$

where (7) requires profit be maximized, (8) equates quantities demanded and supplied in the ith input market, and (9) insures the elimination of each firm's profit in the long run.

The solution of (7)–(9) gives the profit-maximizing rate of output for each firm $q = q^*(p)$, the equilibrium price of the ith input $w^i = w^{i*}(p)$, and the number of firms composing the industry $N = N^*(p)$.[3] These three equations fully characterize the industry's long-run supply as a function of the price of output. The latter is formulated as an exogenous variable because it is clear that in this context it varies directly with the demand for output. Thus, the response of the number of firms to a change in demand can be determined by differentiating (7)–(9), evaluated at equilibrium, with respect to p.

$$\begin{bmatrix} -C_{qq} & -C_{qi} & 0 \\ Nx_q^i & Nx_i^i - g_i & x^i \\ p - C_q & -C_i & 0 \end{bmatrix} \times \begin{bmatrix} dq^*/dp \\ dw^{i*}/dp \\ dN^*/dp \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -q \end{bmatrix} \tag{10}$$

Solving by Cramer's Rule and making substitutions from the expressions in (1)–(7),

$$\frac{dN^*}{dp} = \frac{N(x^i)^2\left[\eta^i - (\eta^i)^2\right] + q^2C_{qq}(Nx_i^i - g_i)}{-qC_{qq}(x^i)^2} \tag{11}$$

Hence, $dN^*/dp > 0$ when $\eta^i \leqq 0$ and $\eta^i \geqq 1$, but $dN^*/dp \gtreqless 0$ as $0 < \eta^i < 1$, so that the expected positive relationship between demand and the number of firms is called into question when the ith input is normal.

II. Changes in Demand and the Firm's Optimal Scale

To show the relationship between changes in demand and the optimal scale of a firm, Cramer's Rule is again employed to solve (10) for

$$\frac{dq^*}{dp} = \frac{(1 - \eta^i)}{C_{qq}} \tag{12}$$

which indicates that optimal scale is inversely related to demand when the ith input is superior and directly related when normal and inferior. Two additional cases should be noted: $\eta^i = 1$ where scale remains constant and $\eta^i = 0$ where scale varies directly.[4] Thus, when the demand for output increases and x^i is superior, the reduced scale requires the entry of new firms to accommodate the increased demand. On the other hand, the larger scale for the cases of normality and inferiority raises the possibility that fewer firms could supply more output.[5] If so, losses must result from the increased demand when exit is prevented.

---

[3] Although the endogenous variables are actually functions of p and w^j $(\forall j \neq i)$, the terms w^j are ignored in the functional notation since no operations involving them appear in the text.

III. Changes in Demand and Profit when Entry and Exit are Prevented

To show the relationship between changes in demand and the profit of existing firms when the number of firms is held constant, the profit function, evaluated at equilibrium,

$$(13)\quad \pi^*(p) = pq^*(p) - C(q^*(p), w^1, \ldots, w^{i*}(p), \ldots, w^n)\cdot$$

is differentiated with respect to p,

$$(14)\quad \left.\frac{d\pi^*}{dp}\right|_{dN=0} = q^* - C_i \left.\frac{dw^{i*}}{dp}\right|_{dN=0}$$

Its sign obviously depends on that of $(dw^{i*}/dp)_{dN=0}$. If dN^*/dp is set equal to zero and the last equation in (10) is omitted, Cramer's Rule can be used to solve for this derivative:[6]

$$(15)\quad \left.\frac{dw^{i*}}{dp}\right|_{dN=0} = \frac{Nx_q^i}{-C_{qq}(Nx_i^i - g_i) + C_{qi}Nx_q^i}$$

Then substituting (15) into (14) and making further substitutions from the relations and definitions in (1)–(6),

$$(16)\quad \left.\frac{d\pi^*}{dp}\right|_{dN=0} = \frac{N(x^i)^2\left[\eta^i - (\eta^i)^2\right] + q^2C_{qq}(Nx_i^i - g_i)}{qC_{qq}(Nx_i^i - g_i) - qN(x_q^i)^2}$$

Since the denominator is negative, the sign of (16) depends on η^i. In fact, the numerator is identical to that of dN^*/dp in (11). Thus, the condition on η^i that determines the response of the number of firms to a change in demand is the same as that which determines the response of profit when the number of firms is held constant. When x^i is superior or inferior, profit would result so entry occurs. On the other hand, when x^i is normal, loss and exit must also be included as possible outcomes.

Holding the number of firms constant so that marginal cost no longer need equal average cost, it is clear that profit will arise from the increase in demand when x^i is inferior. Reducing the demand for x^i, the increase in the industry's output will lower w^i, as is apparent from (15) when it is noted that the denominator is positive since $C_{qi} = x_q^i$ and the numerator is negative only when $\eta^i < 0$. A lower w^i implies a higher marginal cost (since $C_{qi} = x_q^i < 0$) and a lower average cost. Consequently, average revenue rises, average cost falls, and profit results. In long-run equilibrium the number of firms must increase. The larger scale of each firm is not adequate to meet the greater demand for output without entry.

When the input is superior, the consequent increase in its demand raises w^i, average cost, and marginal cost ($C_{qi}=x_q^i>0$), but average revenue rises more than average cost so that profit ensues. Again, the number of firms must increase. Indeed, the smaller scale of each firm requires entry to satisfy the greater demand.

---

[4] It should be noted that where the production function is (globally or locally) homothetic at the point of minimum average cost, $\eta^j = 1 (j = 1, \ldots, n)$ so that $dq^*/dp = 0$, and the phenomenon under study does not occur. See Eugene Silberberg, p. 739, for a discussion of homotheticity and scale adjustments.

[5] There is an important literature dealing with the competitive firm's adjustment of scale to changes in input prices: Ferguson and Thomas Saving; Lowell Bassett and Thomas Borcherding; Richard Portes; S. Charles Maurice (1971, 1972); and Silberberg. The relationship proved by these papers is that $\partial q^*/\partial w^i \gtreqless 0$ as $\eta^i \gtreqless 1$. From (10), $dw^{i*}/dp > 0$ implies in the case of inferiority that although each firm is using less of the ith input at its larger scale, the entry of new firms has increased the input's demand and price.

[6] This derivative differs from its counterpart described in fn. 5. In the latter case production occurs where average and marginal costs are equal so that the change in w^i reflects adjustments in scale and numbers. In the former case marginal and average costs are not necessarily equal so that the change in w^i reflects only the existing firms' alteration in the demand for x^i to maximize profits. This is an extension of D. V. T. Bear's discussion of the relationship between the price of output and the competitive firm's demand for an input. He found that $\partial x^i/\partial p \gtreqless 0$ as $\eta^i \gtreqless 0$.

When the input is normal, the rise in w^i causes both average and marginal cost to increase, but the rise in average cost may exceed the rise in average revenue so that loss as well as profit is possible. Thus, the increased scale for a normal x^i raises the possibility that fewer firms could supply more output and that when exit is prevented, losses would result. Consequently, it cannot be claimed that an increase in demand will always increase profit and cause the entry of new firms.

REFERENCES

L. R. Bassett and T. E. Borcherding, "The Firm, the Industry, and the Long-Run Demand for Factors of Production," *Can. J. Econ.*, Feb. 1970, *3*, 140–44.

D. V. T. Bear, "Inferior Inputs and the Firm," *J. Polit. Econ.*, June 1965, *73*, 287–89.

W. E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *J. Polit Econ.*, May/June 1971, *79*, 481–507.

C. E. Ferguson and John P. Gould, *Microeconomic Theory*, Homewood 1975.

______ **and T. R. Saving,** "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry" *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

G. Hanoch, "The Elasticity of Scale and the Shape of Average Costs," *Amer. Econ. Rev.*, June 1975, *65*, 492–97.

Edwin Mansfield, *Microeconomics: Theory and Applications*, 2d ed., New York 1975.

S. C. Maurice, "Factor-Price Changes, Profit, and Long-Run Equilibrium," *Western Econ. J.*, Mar. 1971, *9*, 64–77.

______, "Long-Run Factor Demand in a Perfectly Competitive Industry," *J. Polit. Econ.*, Nov./Dec. 1972, *80*, 1271–79.

R. D. Portes, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry: An Alternative Approach," *Amer. Econ. Rev.*, June 1971, *61*, 430–34.

E. Silberberg, "The Theory of the Firm in 'Long-Run' Equilibrium," *Amer. Econ. Rev.*, Sept. 1974, *64*, 734–41.

J. Viner, "Cost Curves and Supply Curves," *Z. Nationalökon.*, Sept. 1931, *3*, 23–46.

Tariffs vs. Quotas under Uncertainty: An Extension

By LESLIE YOUNG*

In a recent paper in this *Review*, Partha Dasgupta and Joseph Stiglitz (hereafter D-S) consider whether a pure tariff or a pure quota is better for raising revenue when there is uncertainty about the foreign price of a good and about its domestic demand and supply functions. They conclude that "...a pure tariff is unambiguously superior to a pure quota in generating a given expected level of government revenue" (p. 979). In arriving at this conclusion, the authors use some approximations which they recognize to be valid only when the revenue to be raised—and hence the tariff required—is small. In this note it is shown that, even when this condition is satisfied, tariffs can be inferior to quotas—unless the tariff rate t^* required under a pure tariff is small compared to the degree of uncertainty in the foreign price $\tilde{p}$. If this assumption is reversed, and the degree of uncertainty in the shift parameters $\tilde{\alpha}$ and $\tilde{\gamma}$ in the domestic demand and supply functions is small, then a pure quota can be superior to a pure tariff.

The theorem below gives the ranking of tariffs and quotas when the expected revenue to be raised is small. Let $\bar{p}$ be the expected foreign price and σ_p, σ_α, and σ_γ be the standard deviations of $\tilde{p}$, $\tilde{\alpha}$, and $\tilde{\gamma}$. Let t_0 be the *expected prohibitive tariff*, that is, the tariff rate which ensures that imports are zero if $\tilde{p}$, $\tilde{\alpha}$, and $\tilde{\gamma}$ take their expected values.

THEOREM 1: *In the D-S model with the expected net surplus criterion*: (a) *If* $2t^{*2} > t_0 > \sigma_p/\bar{p} > 0$ *and* σ_α *and* σ_γ *are sufficiently small compared to* σ_p *then a pure quota is superior to a pure tariff in raising a given expected revenue.* (b) *If* $2t^{*2} < t_0 < \sigma_p/\bar{p}$ *then a pure tariff is superior to a pure quota.*

This result seems intuitively more reasonable than the D-S conclusion—which the authors themselves had found surprising. A similar conclusion holds with the max-min criterion. I shall give an intuitive explanation of my results based on the observation that the tariff required under a pure tariff regime in order to raise a given level of expected revenue is higher than the implicit tariff in a pure quota regime.

I. The D-S Model and Conclusions

In the D-S model, the domestic demand and supply schedules are

$$q = \tilde{\alpha} - \beta D; \quad q = \tilde{\gamma} + \delta S$$

where q is domestic price, D and S are the quantities demanded and supplied, $\tilde{\alpha}$ and $\tilde{\gamma}$ are random variables with fixed known distributions, and β and δ are positive constants. The government seeks to maximize the expected sum of consumer's and producer's surplus:

$$W = E\left[(\tilde{\alpha}^2 - \tilde{q}^2)/2\beta - (\tilde{q}^2 - \tilde{\gamma}^2)/2\delta - \tilde{p}\tilde{I}\right] \tag{1}$$

where $\tilde{I}$ is the equilibrium level of imports:

$$\tilde{I} = \frac{\tilde{\alpha}}{\beta} + \frac{\tilde{\gamma}}{\delta} - \tilde{q}\frac{(\beta+\delta)}{\beta\delta} \tag{2}$$

Dasgupta and Stiglitz show that ranking the policies according to W is equivalent to ranking them according to

$$Z \equiv E[\tilde{p}\tilde{q}] - E[\tilde{q}^2]/2 \tag{3}$$

The government is constrained to raise a fixed expected revenue R through tariffs (or through auctioning off quotas), i.e.,

$$E[(\tilde{q} - \tilde{p})\tilde{I}] = R \tag{4}$$

---

*Senior lecturer in economics, University of Canterbury, New Zealand. I am grateful to Edward Tower and to a referee for helpful comments.

In the case of a pure ad valorem tariff $t, \tilde{q}=\tilde{p}(1+t)$. Substituting (2) into (4), D-S show that t satisfies the equation:

$$t^2-\left[\frac{\bar{p}(\bar{\alpha}\delta+\bar{\gamma}\beta)}{(\beta+\delta)E[\tilde{p}^2]}-1\right]t + \frac{R\beta\delta}{(\beta+\delta)E[\tilde{p}^2]}=0$$

(A random variable with a bar denotes its expected value.) Dasgupta and Stiglitz assert that if second and higher powers of R are neglected then an approximate solution to this equation is given by

$$t^* \simeq \frac{R\beta\delta}{\bar{p}(\bar{\alpha}\delta+\bar{\gamma}\beta)-(\beta+\delta)E[\tilde{p}^2]} \tag{5}$$

This implies that

$$\tilde{q}=\left[1+\frac{R\beta\delta}{\bar{p}(\bar{\alpha}\delta+\bar{\gamma}\beta)-(\beta+\delta)E[\tilde{p}^2]}\right]\tilde{p} \tag{6}$$

They assert that, substituting from (6) into (3), one obtains (for small R):

$$Z_{tariff}=E[\tilde{p}^2]/2 \tag{7}$$

In fact, this operation yields

$$Z_{tariff}=E[\tilde{p}^2][(1+t^*)-(1+t^*)^2/2]$$

or

$$Z_{tariff}=E[\tilde{p}^2](1-t^{*2})/2 \tag{8}$$

Notice that D-S have dropped the term $-(1/2)E[\tilde{p}^2]t^{*2}$. The consequences of this will be assessed presently.

In the case of a pure quota, D-S show that the quota I^* required to raise the revenue R satisfies the quadratic equation:

$$I^2-\left[\frac{\bar{\alpha}\delta+\bar{\gamma}\beta-\bar{p}(\beta+\delta)}{\beta\delta}\right]I+\frac{R(\beta+\delta)}{\beta\delta}=0$$

and obtain as an approximation to the relevant solution:

$$I^* \simeq \frac{\bar{\alpha}\delta+\bar{\gamma}\beta-\bar{p}(\beta+\delta)}{\beta\delta} - \frac{R(\beta+\delta)}{\bar{\alpha}\delta+\bar{\gamma}\beta-\bar{p}(\beta+\delta)} \tag{9}$$

They then infer that the equilibrium price under a pure quota is

$$\tilde{q} \simeq \bar{p}+k+\theta \tag{10}$$

where

$$k \equiv R\beta\delta/[\bar{\alpha}\delta+\bar{\gamma}\beta-\bar{p}(\beta+\delta)] \tag{11}$$

$$\theta \equiv [(\tilde{\alpha}-\bar{\alpha})\delta+(\tilde{\gamma}-\bar{\gamma})\beta]/(\beta+\delta) \tag{12}$$

They assert that, substituting equation (10) into (3), one obtains:

$$Z_{quota}=\tfrac{1}{2}\left[\bar{p}^2-(\sigma_\alpha^2\delta^2+\sigma_\gamma^2\beta^2)/(\beta+\delta)^2\right] \tag{13}$$

In fact, this operation yields

$$Z_{quota}=E[\tilde{p}(\bar{p}+k+\theta)]-\tfrac{1}{2}E[(\bar{p}+k+\theta)^2]$$

or

$$Z_{quota}=\tfrac{1}{2}\left[\bar{p}^2-(\sigma_\alpha^2\delta^2+\sigma_\gamma^2\beta^2)/(\beta+\delta)^2-k^2\right] \tag{14}$$

Thus D-S have dropped the term $-(1/2)k^2$ in Z_{quota}.

According to the expressions (7) and (13) obtained by D-S

$$Z_{tariff}-Z_{quota}=\tfrac{1}{2}\left[\sigma_p^2+(\sigma_\alpha^2\delta^2+\alpha_\gamma^2\beta^2)/(\beta+\delta)^2\right] \tag{15}$$

hence D-S conclude that if σ_p, σ_α, and σ_γ are not all zero, then a pure tariff is unambiguously superior to a pure quota.

II. The Optimal Policy under the Expected Surplus Criterion

The D-S conclusion is not valid without further assumptions because the difference $Z_{tariff}-Z_{quota}$ could be of the same order of magnitude as the two neglected terms $-(1/2)E[\tilde{p}^2]t^{*2}$ and $-(1/2)k^2$.[1] To see this, note that by (8) and (14),

$$(16)\qquad 2(Z_{tariff}-Z_{quota})=\sigma_p^2+E[\theta^2]+k^2-E[\tilde{p}^2]t^{*2}$$

where $E[\theta^2]=(\sigma_\alpha^2\delta^2+\sigma_\gamma^2\beta^2)/(\beta+\delta)^2$. To evaluate the importance of the neglected term

$$(17)\qquad k^2-\bar{p}^2t^{*2}-\sigma_p^2t^{*2}$$

rewrite (5) and (11) as

$$(18)\qquad t^*=\frac{aR/\bar{p}}{q_0-\bar{p}-\sigma_p^2/\bar{p}}$$

$$(19)\qquad k=\frac{aR}{q_0-\bar{p}}$$

where $a\equiv\beta\delta/(\beta+\delta)$ and $q_0\equiv(\bar{\alpha}\delta+\bar{\gamma}\beta)/(\beta+\delta)$. Thus q_0 is the domestic equilibrium price if imports are zero and $\tilde{\alpha}$ and $\tilde{\gamma}$ take their expected values.

From (18) and (19) it can be seen that if $\sigma_p>0$ then $k<\bar{p}t^*$ and the neglected term (17) is negative.[2] It is not immediately apparent whether the D-S conclusion remains valid because the size of the neglected term (17) depends on σ_p^2—and the term σ_p^2 appears independently in the D-S criterion function (15). However, with some elementary manipulations (see Appendix A), we have

$$(20)\qquad 2(Z_{tariff}-Z_{quota})=E[\theta^2]+\sigma_p^2[1-2\bar{p}t^{*2}/(q_0-\bar{p})]+t^{*2}\sigma_p^2\left[-1+\sigma_p^2/(q_0-\bar{p})^2\right]$$

Assuming that $\sigma_p>0$ the last two terms in this expression are negative if

$$(21)\qquad 2t^{*2}>(q_0-\bar{p})/\bar{p}>\sigma_p/\bar{p}$$

They are positive if

$$(22)\qquad 2t^{*2}<(q_0-\bar{p})/\bar{p}<\sigma_p/\bar{p}$$

Thus the D-S conclusion is valid if condition (22) holds. However, if condition (21) holds then the last two terms in (20) are negative. Noting that $E[\theta^2]=[\delta^2\sigma_\alpha^2+\beta^2\sigma_\gamma]/(\beta+\delta)^2$, I conclude that if σ_α^2 and σ_γ^2 are sufficiently small relative to σ_p^2, then quotas are superior to tariffs. Theorem 1 is proved.

In Appendix B, I show that it is possible to choose values for the various parameters so that (21) is satisfied *without* violating the conditions under which the approximations for t^* and I^* are valid.

Conditions (21) and (22) have a simple interpretation: $(q_0-\bar{p})/\bar{p}$ is the *expected prohibitive tariff* t_0, that is, the tariff that would lead to zero imports if $\tilde{p}$, $\tilde{a}$, and $\tilde{\gamma}$ took their expected values. The condition $2t^{*2}>t_0$ thus requires that the pure tariff t^* required to raise the given expected revenue R be not too small compared to the expected prohibitive tariff. The condition $t_0>\sigma_p/\bar{p}$ or $q_0-\bar{p}>\sigma_p$ requires that, on average, the fluctuation in foreign prices be less than the difference between the domestic equilibrium price when imports are zero and the foreign price.

III. The Optimal Policy under the Max-Min Criterion

Dasgupta and Stiglitz also considered the ranking of tariffs and quotas when there is infinite risk aversion with respect to net consumer's surplus, rather than risk neutrality as above. Policies are ranked according to the minimum value of net consumer's surplus over all values of the random variables, i.e.,

---

[1] Dasgupta and Stiglitz state that the approximation (5) is obtained by neglecting second and higher powers of R and that their expression (7) holds for small R. Thus they seem to suggest that if the approximation (5) is used, then the term $-(1/2)E[\tilde{p}^2]t^{*2}$ *ought* to be neglected. In Appendix B I shall examine the conditions under which approximations (5) and (9) are valid and show that these conditions need *not* imply that the neglected terms are small compared to $Z_{tariff}-Z_{quota}$.

[2] I assume that $q_0>\bar{p}+\sigma_p^2/\bar{p}$ so that $k>0$ and $t^*>0$. This assumption ensures that the tariff or quota is binding. Dasgupta and Stiglitz make a similar assumption in their fn. 8.

$$\hat{W} = min\left[(\tilde{\alpha}^2 - \tilde{q}^2)/2\beta - (\tilde{q}^2 - \tilde{\gamma}^2)/2\delta - \tilde{p}\tilde{I}\right] \tag{23}$$

Routine calculations (see Appendix C) yield

$$\hat{W}_{tariff} - \tilde{W}_{quota} = \frac{\beta+\delta}{2\beta\delta} \times \left[p_{max} - (\bar{p} + k + \theta_{max})\right]^2 - \frac{\beta+\delta}{2\beta\delta}\left[p_{max}t^*\right]^2 \tag{24}$$

where p_{max} and θ_{max} are the largest values of $\tilde{p}$ and θ, respectively. However, in ranking the policies, D-S ignore the last term so again they conclude that tariffs are unambiguously superior to quotas. When this term is considered, we have

THEOREM 2: *Under the maxmin criterion, quotas are superior to tariffs in raising a given expected revenue if and only if*:

$$|p_{max} - (\bar{p} + k + \theta_{max})| < p_{max}t^* \tag{25}$$

Condition (25) states that the maximum foreign price should differ from the maximum domestic price under a pure quota by less than the maximum tariff payment per unit of the commodity under a pure tariff. This condition is satisfied in nontrivial cases. For example, since (18) and (19) imply that $k < \bar{p}t^*$, we can choose θ_{max} so that $\bar{p} + k + \theta_{max} < p_{max} + p_{max}t^*$. If, in addition, p_{max} is chosen so that $p_{max} - \bar{p} < k + \theta_{max}$, then $p_{max}t^* > \bar{p} + k + \theta_{max} - p_{max} = |p_{max} - (\bar{p} + k + \theta_{max})|$. Thus if $p_{max} - \bar{p}$ and θ_{max} are sufficiently small compared to the maximum tariff payments under a pure tariff, then quotas are superior to tariffs.

IV. Concluding Remarks

I have brought out explicitly the assumptions underlying the D-S conclusion that tariffs are superior to quotas as revenue-raising devices under uncertainty. Moreover, for both the expected net surplus and the max-min criteria, I have stated conditions under which the reverse ranking holds. For both welfare criteria, the condition is that the pure tariff rate be large compared to the degree of uncertainty in foreign prices and in the domestic demand and supply functions—different measures of the degree of uncertainty being used with the two criteria. I now give an intuitive interpretation of these results.

The welfare criterion (1) can be written as $B - C - M$ where $B \equiv E[(\tilde{\alpha}^2 - \tilde{q}^2)/2\beta)]$, $C \equiv E[(\tilde{q}^2 - \tilde{\gamma}^2)/2\delta]$, and $M \equiv E[\tilde{p}\tilde{I}]$. The expressions $B - C$ and M embody two sets of factors affecting the ranking of tariffs and quotas: a pure tariff is inferior on one count and superior on the other. Under a pure tariff, imports are negatively correlated with the foreign price. Therefore,

$$E[\tilde{p}\tilde{I}] < \bar{p}E[\tilde{I}] \tag{26}$$

That is, the expected cost of imports under a pure tariff t is less than the expected cost if the expected imports were imposed as a pure quota. With the import function (2):

$$M_{quota} - M_{tariff} = \{E[\tilde{p}^2] - \bar{p}^2\} \times (1+t)/a = \sigma_p^2(1+t)/a > 0$$

Thus tariffs are superior to quotas in terms of expected import costs M. Clearly they are also superior because they permit imports to respond to changes in the parameters $\tilde{\alpha}$ and $\tilde{\gamma}$ in the domestic demand and supply functions.

Leaving the last-mentioned considerations aside, let t' be the implicit tariff from auctioning off a given quota if $\tilde{\alpha}$ and $\tilde{\gamma}$ take their expected values. If t' were set as a pure tariff then:

$$(B-C)_{tariff} - (B-C)_{quota} = \{-E[\tilde{p}^2] + \bar{p}^2\} \times (1+t')^2/2a = -\sigma_n^2(1+t')^2/2a < 0$$

Thus tariffs are inferior to quotas in terms of the "risk-aversion" effects arising because the consumer's gross benefit B is a concave function of q and the domestic cost curve C is a convex function of $\tilde{q}$. Since $\sigma_p^2(1+t')/a - \sigma_p^2(1+t')^2/2a = \sigma_p^2(1-t'^2)/2a > 0$ for $t' < 1$ we could conclude that a pure tariff is superior to a pure quota in raising a given revenue, provided that the revenue is small (so $t' < 1$) *and* provided that the pure tariff rate t^* which generates the given expected revenue equals the implicit tariff rate t' realized by auctioning off the required quota. The neglect of the terms k^2 and $E[\tilde{p}^2]t^{*2}$ by D-S is equivalent to making this last assumption.

In fact, $t^* > k/\bar{p} = t'$ because the expected cost M of given expected imports (and hence the expected tariff revenue for a given tariff rate) is less under a pure tariff than under a pure quota (see equation (26)). The larger tariff required under a pure tariff regime to raise a given expected revenue reduces the term $(B-C)_{tariff}$ since it increases both the mean and the standard deviation of domestic prices. The revenue constraint thus implies an indirect connection between the two sets of factors, embodied in $B-C$ and M, which were noted separately by D-S. The fact that a pure tariff is superior in terms of expected import costs gives an *additional* reason why they are inferior in terms of risk-aversion effects. The neglect by D-S of the terms $k^2 - E[\tilde{p}^2]t^{*2}$ corresponds to neglect of this indirect connection. Taking this connection into account, we have the ranking given by Theorem 1. The rationale of Theorem 2 is similar.

APPENDIX A: DERIVATION OF EQUATION (20)

By (18) and (19)

$$k^2 - \bar{p}^2 t^{*2} = \bar{p}^2 t^{*2}\left[\frac{\left(q_0 - \bar{p} - \sigma_p^2/\bar{p}\right)^2}{\left(q_0 - \bar{p}\right)^2} - 1\right]$$

$$= \frac{\bar{p}^2 t^{*2}}{\left(q_0 - \bar{p}\right)^2}\left[-2(q_0 - \bar{p})\sigma_p^2/\bar{p} + \sigma_p^4/\bar{p}^2\right]$$

Therefore

$$k^2 - \bar{p}^2 t^{*2} + \sigma_p^2(1 - t^{*2}) = \sigma_p^2\left[1 - 2\bar{p}t^{*2}/(q_0 - \bar{p})\right] + t^{*2}\sigma_p^2\left[-1 + \sigma_p^2/(q_0 - \bar{p})^2\right]$$

Substitution in (16) then yields (20).

APPENDIX B

I show that the condition (21) for quotas to be superior to tariffs is consistent with the conditions for the approximations for t^* and I^* to be valid. These approximations are obtained as follows. The solution to the quadratic equation $x^2 - bx + c$ is $(1/2)[b \pm b(1 - 4c/b^2)^{(1/2)}]$. If $1 \gg 4c/b^2$, i.e., $4c/b^2$ is small compared to 1, then second and higher powers of $4c/b^2$ can be neglected and we have the approximations $x \simeq c/b$ or $x \simeq b - c/b$. In (5) $t^* \simeq c_1/b_1$ where $c_1 = aR/(\bar{p}^2 + \sigma_p^2)$ and $b_1 = [\bar{p}(q_0 - \bar{p}) - \sigma_p^2]/(\bar{p}^2 + \sigma_p^2)$. In (9) $I^* \simeq b_2 - c_2/b_2$ where $c_2 = R/a$ and $b_2 = (q_0 - \bar{p})/a$.

For any $\bar{p}$ we can choose r and $q_0 > \bar{p}$ so that $1 \gg 2\bar{p}/r = 4r/(q_0 - \bar{p})$. Setting $R = r(q_0 - \bar{p})/a$ we then have

$$\text{(A1)} \qquad 1 \gg \frac{2\bar{p}(q_0 - \bar{p})}{aR} = \frac{4aR}{(q_0 - \bar{p})^2}$$

Equation (A1) implies that $4c_2/b_2^2 \ll 1$ so approximation (9) for I^* is valid. Moreover, for a sufficiently small σ_p such that $0 < \sigma_p < q_0 - \bar{p}$ we clearly have

$$1 \gg \frac{4aR\left(\bar{p}^2 + \sigma_p^2\right)}{\left[\bar{p}(q_0 - \bar{p}) - \sigma_p^2\right]^2} = \frac{4c_1}{b_1^2}$$

so approximation (5) for t^* is valid. Finally, from (A1) we have

$$\frac{2a^2R^2}{\left[\bar{p}(q_0 - \bar{p}) - \sigma_p^2\right]^2} > \frac{2a^2R^2}{\bar{p}^2(q_0 - \bar{p})^2}$$

$$= \frac{q_0 - \bar{p}}{\bar{p}} > \frac{\sigma_p}{\bar{p}}$$

Thus the condition (21) for quotas to be superior to tariffs is satisfied.

APPENDIX C: DERIVATION OF EQUATION (24)

Recalling (12) we can rewrite (2) as

$$a\tilde{I} = q_0 + \theta - \tilde{q} \tag{A2}$$

where $a = \beta\delta/(\beta+\delta)$. Let $A \equiv (\delta\alpha_{min}^2 + \beta\gamma_{min}^2)/(\beta+\delta)$. Then substituting from (A2) into (23) we have

$$2a\hat{W}_{tariff} = A - q_{max}^2 - 2p_{max}(q_0 + \theta_{max} - q_{max}) \tag{A3}$$

Since $q_{max} = p_{max}(1+t^*)$, (A3) simplifies to

$$2a\hat{W}_{tariff} = A + p_{max}^2 - 2p_{max}(q_0 + \theta_{max}) - (p_{max}t^*)^2 \tag{A4}$$

By (9) and (10) we have

$$aI^* \simeq q_0 - \bar{p} - k \tag{A5}$$

$$\tilde{q} = \bar{p} + k + \theta \tag{A6}$$

Therefore, substituting (A5) and (A6) into (23) we have

$$2a\hat{W}_{quota} = A - (\bar{p} + k + \theta_{max})^2 - 2p_{max}(q_0 - \bar{p} - k) \tag{A7}$$

By (A3) and (A7)

$$2a(\hat{W}_{tariff} - \hat{W}_{quota}) = p_{max}^2 + (\bar{p} + k + \theta_{max})^2 - 2p_{max}(\bar{p} + k + \theta_{max}) - (p_{max}t^*)^2$$

$$= [p_{max} - (\bar{p} + k + \theta_{max})]^2 - (p_{max}t^*)^2$$

Equation (24) now follows immediately.

REFERENCES

P. Dasgupta and J. E. Stiglitz, "Tariffs vs. Quotas as Revenue-Raising Devices under Uncertainty," *Amer. Econ. Rev.*, Dec. 1977, *67*, 976–81.

Further Results on Input Choices under Uncertain Demand

By SATYA P. DAS*

In a recent article in this *Review*, Duncan Holthausen examined whether or not the well-known correspondence or duality between profit maximization and cost minimization extends to the presence of uncertain demand. He showed that this duality holds for a quantity-setting firm but not for a "price-setting" firm.[1] The purpose of this paper is to complement and to partially correct some of Holthausen's results. My findings can be briefly stated as follows: First, there can be three modes of behavior for a firm in the presence of uncertainty: quantity setting, price setting, and quantity price setting (see Hayne Leland for example). Whereas Holthausen considered the first two, the third case was left out, somewhat surprisingly. The case of a quantity-price-setting firm is analyzed in this paper, and it turns out that the duality between (expected utility of) profit maximization and cost minimization holds.

I show this in the two-input case. But it is generalizable to the case of any number of inputs for a quantity-price-setting firm as well as for a quantity-setting firm. In view of the fact that all inputs are chosen *ex ante* in the case of these two types of firms, the following theorem is then available.

THEOREM 1: *Under demand uncertainty, if all inputs are chosen ex ante, then the duality between (expected utility of) profit maximization and cost minimization holds.*

When some inputs are chosen *ex post* and some *ex ante*, the output becomes random, and hence the concept of "total cost" becomes random too. Thus what is relevant here is the expected utility of cost minimization (as examined in Theorem 2 below) or simply expected cost minimization (as Holthausen has examined). For a price-setting firm under demand uncertainty, Holthausen has shown in his two-input framework that when one input is chosen *ex post*, (i) the expected utility of profit maximization does not imply expected cost minimization (or in other words, in a rough sense, the duality between profit maximization and cost minimization does not hold); (ii) moreover, under risk aversion the optimal amount of the input chosen *ex ante* under expected utility of profit maximization is less than that under expected cost minimization.[2] Holthausen's treatment of such a price-setting firm, however, seems somewhat unsatisfactory. For he compared input choice under expected utility of profit maximization to that under expected cost minimization (rather than expected utility of cost minimization). This means that whereas the firm can be risk averse, risk neutral or risk preferring with respect to its profit, it can only be risk neutral with respect to its cost! A more appropriate way, I believe, is to allow a non-linear utility function for both profit and cost, especially the one given by risk-averse behavior.

THEOREM 2: *For a price-setting firm under demand uncertainty, expected utility of profit maximization does not imply expected utility of cost minimization; and the optimal amount of the input chosen ex ante (under expected utility of profit maximization) is less than that*

---

*University of Wisconsin-Milwaukee. I am thankful to George H. Borts, Yoshio Niho, and an anonymous referee for comments which led to a marked improvement of this paper.

[1] Strictly speaking, along with a quantity-setting firm, Holthausen considers a partial quantity-price-setting firm rather than simply a price-setting firm, because in his Sections III–VII, the price as well as the capital stock are the *ex ante* variables.

[2] Note that cost minimization here assumes output not to be constant but equal to the quantity demanded.

under expected utility of cost minimization if the two inputs are complementary in production and if the measure of the firm's absolute risk aversion is constant.

Thus, Holthausen's result (i) essentially holds whereas his result (ii) holds with some additional restrictions. Moreover, the first part of Theorem 2 easily extends to the case of any number of inputs whereas its second part does not.

When some inputs are chosen *ex post* and some *ex ante*, the familiar marginality conditions do not hold either for (expected utility of) profit maximization or for (expected utility of) cost minimization; however it may be of some interest to know if there is any special case where they do. This is discussed in Section III where a three-input model is developed. One input is assumed to be used *ex post* and the other two are assumed to be used *ex ante.*

THEOREM 3: *If the production function is separable*[3] *in the input used ex post, then the expected utility of profit maximization and the expected utility of cost minimization both imply that the ratio of the marginal products of the inputs used ex ante equals the ratio of their prices.*

Theorem 3 can be easily extended to the case of any number of inputs used *ex ante.* In the following sections, Theorems 1–3 are proved.

I. Proof of Theorem 1

As Holthausen has already considered the quantity-setting firm which complies with Theorem 1, I only examine the quantity-price-setting firm.

The features of such a model are well known (see Edwin Mills and Leland) and therefore only a brief description will be given. Let $q=h(p,u)$ be the demand function for the firm's product where q,p, and u denote the quantity demanded, price, and the random term. It is assumed that $h_p<0$ and, without loss of generality, $h_u>0$, where the subscripts denote the partials. Let α and β be the lower and upper limits of u. Let $X=f(K,L)$ be the production function, where X, K, and L are the output and the amounts of capital and labor, respectively. Because p and X are determined before the actual value of u is known, the quantity demanded may differ from the level of output. For any given p and X, let us define u_0 such that $X=h(p,u_0)$. Thus if $u \geqslant u_0$, the quantity sold is X and if $u \leqslant u_0$, it is $h(p,u)$. Then the firm's profit, π, is given by

$$\pi_1=pf(K,L)-wL-rK \quad \text{for } u_0 \leqslant u \leqslant \beta$$

$$\pi_2=ph(p,u)-wL-rK \quad \text{for } \alpha \leqslant u \leqslant u_0$$

where w and r are the given prices of labor and capital, respectively.

Thus the expected utility of profit is given by

$$(1)\quad \int_{\alpha}^{u_0} U[ph(p,u)-wL-rK]\phi(u)du + U[pf(K,L)-wL-rK]\int_{u_0}^{\beta}\phi(u)du$$

where $\phi(u)$ is the density function of u. The firm's objective is to maximize (1) with respect to K, L, and p. The first-order conditions are[4]

$$(2)\quad pf_K U'(\pi_1)\int_{u_0}^{\beta}\phi(u)\,du - rE[U'(\pi)]=0$$

$$(3)\quad pf_L U'(\pi_1)\int_{u_0}^{\beta}\phi(u)\,du - wE[U'(\pi)]=0$$

$$(4)\quad \int_{\alpha}^{u_0} U'(\pi_2)(h+ph_p)\phi(u)\,du + U'(\pi_1)f\int_{u_0}^{\beta}\phi(u)\,du=0$$

where f_K and f_L are the marginal products.

---

[3] Wassily Leontief's paper is one of the earlier works on separable production functions. On the demand side, there exists a large body of literature on separable utility functions, for example, Steven Goldman and Hirofumi Uzawa.

[4] Note that u_0 is a function of p, K, and L. But the terms involving its partials cancel out in each of the first-order conditions.

From (2) and (3), it readily follows that $f_K/f_L=r/w$. Thus a quantity-price-setting firm minimizes its cost. This result is easily generalizable for n factors.

II. Proof of Theorem 2

Theorem 2 involves a comparison between the firm's expected utility of profit maximization and its expected utility of cost minimization. Let us consider the latter first.

Inverting the production function, $X=f(K,L)$, let us define the labor requirement function as $L=H(X,K)$. Some of the properties of this function are

$$\frac{\partial H}{\partial X}=\frac{1}{f_L}>0$$

$$\frac{\partial H}{\partial K}=-\frac{f_K}{f_L}<0$$

$$\frac{\partial^2 H}{\partial X\cdot\partial K}=\frac{f_{LL}f_K-f_{LK}f_L}{f_L^3}<0 \quad \text{if } f_{LK}>0$$

that is, if labor and capital are complementary to each other.

Let us now define the negative of the total cost as

$$\begin{aligned} C &\equiv -wL-rK \\ &\equiv -wH(X,K)-rK \\ &= -wH(h(p,u),K)-rK \end{aligned}$$

because for a price-setting firm, the quantity produced is always equal to the quantity demanded. This implies, as in Holthausen's paper, that labor is being used *ex post.* Clearly, for any given p and K, C is random, and the equivalent problem for cost minimization under uncertainty will be to maximize $E[U(C)]$, where E is the expectation operator. Equating its partial with respect to K, to zero, we get

$$\text{(5)} \quad E[U'(C)(-w\partial H/\partial K-r)]=0$$

let K^0 be the solution to (5) at a given p.

Now the profit π of the price-setting firm is equal to $R+C$, where $R\equiv$the total revenue$\equiv ph(p,u)$. Then the objective function of the firm can be written as

$$\underset{p,K}{\text{Max}}\, E[U(R+C)]$$

The first-order conditions are[5]

$$\text{(6a)} \quad E[U'(\pi)\{h+h_p(p-w\partial H/\partial X)\}]=0$$

$$\text{(6b)} \quad E[U'(\pi)(-w\partial H/\partial K-r)]=0$$

Let K^* be the optimal amount of capital determined from equations (6a) and (6b). The question is how K^0 and K^* are related. A close look at (5) and (6b) suggests that $K^0\neq K^*$ or, in other words, as Holthausen contends input choices under expected utility of profit maximization are not consistent with cost minimization. This proves the first part of Theorem 2.

To compare the magnitude of K^* relative to K^0, let us evaluate the left-hand side of (6b) at K^0.[6] By Taylor's approximation, $U'(\pi)=u'(C)+RU''(C)$ Substituting this in (6b) and evaluating (6b) at K^0, we get

(7) left-hand side of (6b)

$$\begin{aligned} &=E[RU''(C)(-w\partial H/\partial K-r)] \\ &=E[R]E[U''(C)(-w\partial H/\partial K-r)] \\ &\quad +\text{cov}[R,U''(C)(-w\partial H/\partial K-r)] \end{aligned}$$

An inspection of the expression (7) suggests that, in general, its sign is indeterminate. However with constant absolute risk aversion and complementarity between inputs,[7] it will be shown below that the expression (7) is negative in sign. Thus $K^*<K^0$, and hence the second part of Theorem 2 is proved.

---

[5] These equations are the same as equations (15) and (16) of Holthausen's paper.

[6] Note that both K^* and K^0 are evaluated at the price determined by (6a) and (6b).

[7] In production functions such as *CES* or Cobb-Douglas, inputs are in fact complementary.

The measure of absolute risk aversion following Kenneth Arrow and John Pratt is given by $-U''(C)/U'(C)$. Thus constant absolute risk aversion implies that $U''(C)/U'(C)$ is constant. Using this as well as equation (5), (7) can be written as

(8)

$$\frac{U''(C)}{U'(C)}E[R]E[U'(C)(-w\partial H/\partial K-r)]$$

$$+\operatorname{cov}[R,U''(C)(-w\partial H/\partial K-r)]$$

$$=\frac{U''(C)}{U'(C)}\operatorname{cov}[R,U'(C)(-w\partial H/\partial K-r)]$$

From risk aversion, as we know, $U''(C)<0$. This, together with the fact that $\partial H/\partial X>0$ implies that $U'(C)$ and u are positively related. Also because $\partial^2 H/\partial X\cdot\partial K<0$ when the inputs are complementary, $(-w\partial H/\partial K-r)$ is positively related to u. Hence $U'(C)(-w\partial H/\partial K-r)$ is positively related to u. Moreover, R is obviously positively related to u. Thus

$$\operatorname{cov}[R,U'(C)(-w\partial H/\partial K-r)]>0$$

This together with U'' being negative implies that expression (8) and hence expression (7) are negative in sign.

III. Proof of Theorem 3

In this section I will show that for a price-setting firm when some inputs are used *ex ante* and some *ex post*, the familiar marginality conditions (that the ratio of marginal products of two inputs be equal to their relative prices) do hold under both expected utility of profit maximization and expected utility of cost minimization, if the production function is separable in the input used *ex post*. This is shown in a three-input model, which can be easily generalized to the case of any number of inputs used *ex ante*.

Let the third factor be denoted by N with p_n as its price. Let us maintain our assumption that labor is the factor used *ex post* in order to adjust output to varying demand. Let our separable production function be of the form

$$X=F[g(K,N),v(L)] \tag{9}$$

and correspondingly the labor requirement function be of the form

$$L=H^*(X,K,N)$$

In view of the discussion in Section II, the minimization of expected utility of cost is given by

$$E[U'(C)(-w\partial H^*/\partial K-r)]=0 \tag{10}$$

$$E[U'(C)(-w\partial H^*/\partial N-p_n)]=0$$

From (9) it follows that[8]

$$\partial H^*/\partial K=-F_g g_K/F_L;\partial H^*/\partial N \tag{11}$$

$$=-F_g g_N/F_L$$

Substituting (11) in (10) and rearranging the terms, we get

$$wE[U'(C)F_g g_K/F_L]=rE[U'(C)] \tag{12}$$

$$wE[U'(C)F_g g_N/F_L]=p_nE[U'(C)] \tag{13}$$

But the left-hand side of (12)

$$=wE\left[\frac{U'(C)F_g g_N}{F_L}(F_g g_K/F_g g_N)\right]$$

$$=(g_K/g_N)wE[U'(C)F_g g_N/F_L]$$

as the g function is independent of L

$$=(g_K/g_N)p_nE[U'(C)],\quad \text{using}(13).$$

Hence

$$rE[U'(C)]=(g_K/g_N)p_nE[U'(C)]$$

$$g_K/g_N=r/p_n \tag{14}$$

---

[8] Note that $F_L=F_v v'$.

In other words, the marginality conditions obtained under certainty do extend to the case of expected utility of cost minimization.

Let us now consider maximization of expected utility of profit. In view of the discussion in Section II, it is clear that the first-order conditions of this maximization with respect to K and N are the same as the conditions (12) and (13) except for $U'(\pi)$ in place of $U'(C)$. Thus equation (14) follows. In other words, the expected utility of profit maximization also implies the marginality conditions.

REFERENCES

Kenneth J. Arrow, *Essays in the Theory of Risk Bearing*, Chicago 1971.

S. M. Goldman and H. Uzawa, "A Note on Separability in Demand Analysis," *Econometrica*, July 1964, *32*, 387–98.

D. M. Holthausen, "Input Choices and Uncertain Demand," *Amer. Econ. Rev.*, Mar. 1976, *66*, 94–103.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

W. W. Leontief, "Introduction to a Theory of Internal Structure of Functional Relationships," *Econometrica*, Oct. 1947, *15*, 361–73.

E. J. Mills, "Uncertainty and Price Theory," *Quart. J. Econ.*, Feb. 1959, *73*, 116–29.

J. W. Pratt "Risk-Aversion in the Small and in the Large," *Econometrica*, Jan./Apr. 1964, *32*, 122–36.

NOTES

NINETY-THIRD ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

Denver, Colorado, September 4–7, 1980

Preliminary Announcement of the Program

Thursday, September 4, 1980

10:00 A.M. Executive Committee Meeting

Friday, September 5, 1980

8:00 A.M. Recent Developments in Economic Theory: Austrian Economics
Presiding: M. Bruce Johnson, University of California-Santa Barbara
Papers: Gerald P. O'Driscoll, Jr. and Mario J. Rizzo, New York University
What is Austrian Economics?
J. Huston McCulloch, Ohio State University
The Moderate Quantity Theory
Roger W. Garrison, Auburn University
Intertemporal Coordination and Macro-Economic Stability
Discussants: Mark Perlman, University of Pittsburgh
T. K. Rymes, Carleton University
Mack Ott, Pennsylvania State University

10:15 A.M. Models for Antitrust*
Presiding: Peter O. Steiner, University of Michigan
Papers: Charles R. Plott, California Institute of Technology
Allocation by Committee: The Distribution of Airport Capacity
Richard J. Gilbert, University of California-Berkeley
Investment Decisions with Scale and Learning Effects: Lessons from the Dupont Case
Elizabeth E. Bailey, Civil Aeronautics Board
Entry Hurdles, Contestability, and Antitrust Policy
Joseph E. Stiglitz, Princeton University
When Can Potential Competition Reduce Welfare?
Discussants: Steven N. Salop, Federal Trade Commission
Paul Joskow, Massachusetts Institute of Technology
Janusz Ordover, New York University

10:15 A.M. Application of Theory to the Economics of the Arts*
Presiding: Virginia Lee Owen, Illinois State University
Papers: Roger McCain, Temple University
Cultivation of Taste: Catastrophe Theory and the Demand for Works of Art
Bruce A. Seaman, Georgia State University
Economic Theory and the Positive Economics of Arts Financing
Leslie Singer, Indiana University, Northwest
Supply Decisions by Professional Artists
Discussants: William J. Baumol, Princeton University and New York University
John Michael Montias, Yale University

10:15 A.M. Inventories and Macroeconomics*
Presiding: Louis J. Maccini, Johns Hopkins University
Papers: Alan S. Blinder, Princeton University
Inventories and the Structure of Macroeconomic Models
Louis J. Maccini, Johns Hopkins University, and Robert Rossana, Federal Reserve Bank of Philadelphia
An Analysis of Investment in Finished Goods Inventories

F. OWEN IRVINE, Wesleyan University and Federal Reserve Board
Merchant Wholesaler Inventory Investment and the Cost of Capital
Discussants: FRANK P. R. BRECHLING, University of Maryland
JOHN CARLSON, Purdue University
MASAHIRO KAWAI, Johns Hopkins University

10:15 A.M. APPLIED MARXIAN ANALYSES OF STAGFLATION*
Presiding: DAVID M. GORDON, New School for Social Research
Papers: MICHELE J. NAPLES, University of Massachusetts-Amherst
Does Industrial Conflict Help Explain the Stagnation of Productivity?
SAMUEL ROSENBERG, University of California-Davis, AND THOMAS E. WEISSKOPF, University of Michigan
Does Social Conflict Help Explain the Acceleration of Inflation?
DAVID M. GORDON, New School for Social Research
Stagnant Productivity and Rampant Inflation—Mysterious Residuals or Hidden Structural Connections?
Discussants: MARTIN N. BAILY, The Brookings Institution
HOWARD SHERMAN, University of California-Riverside

10:15 A.M. RECENT ADVANCES IN ENVIRONMENTAL ECONOMICS (Joint Session with the Association of Environmental and Resource Economists)
Presiding: ALLEN V. KNEESE, Resources for the Future, Inc.
Papers: BRUCE A. FORSTER, University of Guelph
Separability, Functional Structure, and Aggregation for a Class of Environmental Models
CURTIS ANDERSON, University of Wyoming
On the Theory of Supply of Non-Renewable Natural Resources
LARRY EUBANKS, Pennsylvania State University
Household Production and Non-Market Valuations
SHAUL BEN-DAVID, University of New Mexico, THOMAS CROCKER, AND SHELBY GERKING, University of Wyoming
Effects of Environmental Factors on Reported Health Status: Some Empirical Results
Discussants: MAUREEN L. CROPPER, University of Southern California
V. KERRY SMITH, University of North Carolina

10:15 A.M. EMERGING ISSUES IN ECONOMIC DEVELOPMENT OF THE THIRD WORLD (Joint Session with the National Economic Association)
Presiding: RANSFORD PALMER, Howard University
Papers: RAWLE FARLEY, State University of New York-College at Brockport
Management and Economic Development
TETTEH KOFI, University of Notre Dame
Intra-African Trade Expansion as a Strategy in Economic Development
VINCENT MCDONALD, Howard University
Regional Development Institutions in the Economic Growth of Caribbean Countries
BENJAMIN BOBO, University of California-Riverside
Multinational Corporations in the Economic Development of Black Africa
Discussants: KEMPE HOPE, Daemon College
DESMOND JOLLY, University of California-Davis

10:15 A.M. FEDERAL CREDIT ACTIVITY (Joint Session with the American Finance Association)
Presiding: DAVID SMALL AND MARY KAY PLANTES, University of Wisconsin-Madison
Papers: DAVID SMALL AND MARY KAY PLANTES, University of Wisconsin-Madison
Macroeconomic Consequences of Federal Credit Activity
KEVIN VILLANI, U.S. Federal Housing Administration
Pricing the Characteristics of Secondary Mortgage Market Financing Vehicles
RAYMOND LOMBRA AND MICHAEL WASYLENKO, Pennsylvania State University
An Appraisal of Microeconomic and Macroeconomic Rationales for Federal Credit Programs: The Case of the Small Business Administration
KENNETH T. ROSEN, University of California-Berkeley
A Microeconomic Model of the Federal National Mortgage Association
Discussants: FRANK DE LEEUW, U.S. Bureau of Economic Analysis
DONALD HESTER, University of Wisconsin
CRAIG SWAN, University of Minnesota

10:15 A.M. DETERMINANTS OF PHYSICIANS' FEES (Joint Session with the Health Economics Research Organization)
Presiding: JON GABEL, U.S. Department of Health, Education, and Welfare
Papers: JERRY CROMWELL, Health Economics Research, Inc.
Impact of Physician Competition on Fee Setting
LEWIS FREIBERG, Blue Cross/Blue Shield Associations
Quality as a Factor in Physician Fee Dispersion
DONALD YETT, WILLIAM DER, RICHARD ERNST, AND JOEL HAY, University of Southern California
Physician Charges and Reimbursement under Blue Shield-Administered Private and Public Health Insurance
Discussants: EDWARD F. MEEKER, American Medical Association
ROBERT H. LEE, The Urban Institute
AMY TAYLOR, National Center for Health Services Research

10:15 A.M. THE ECONOMIC IMPACT OF AGING POPULATION: EXCHANGE AND GRANT TRANSACTIONS (Joint Session with the Association for the Study of the Grants Economy)
Presiding: KENNETH E. BOULDING, University of Colorado
Papers: LARRY D. SINGELL, University of Colorado
Measuring Inequality Through the Life Cycle
ALICIA H. MUNNELL, Federal Reserve Bank of Boston
The Redistributive Impact of Social Security, Pension Plans, and Tax Incentives to Save
THOMAS F. WILSON, Butler University
Redistributive Flows Within the Financial System Affecting Life Cycle Income
ROBERT L. CLARK, North Carolina State University and Duke University
The Impact and Inflation on the Aged: Exchange and Grant Components
Discussants: (To be announced)

10:15 A.M. PRODUCTION THEORY REVISITED
Presiding: INGRID H. RIMA, Temple University
Papers: TEJ S. SAINI, Bloomsburg State College
The Cobb-Douglas Tide and its Aftermath
EARL O. HEADY, Iowa State University
Theory, Forms, and Applications in Agriculture
LAWRENCE J. LAU, Stanford University
Recent Developments in the Theory of Production: A Survey
Discussants: ROBERT M. SOLOW, Massachusetts Institute of Technology
W. ERWIN DIEWERT, University of British Columbia
ERNST R. BERNDT, Massachusetts Institute of Technology

10:15 A.M. CURRENT ISSUES IN MINIMUM WAGE RESEARCH
Presiding: RONALD S. WARREN, JR., University of Virginia
Papers: CHARLES C. BROWN, CURTIS L. GILROY, AND ANDREW I. KOHEN, Minimum Wage Study Commission
Employment Effects of the Minimum Wage
EDGAR K. BROWNING AND WILLIAM R. JOHNSON, University of Virginia
The Effect of Changing the Minimum Wage on the Net Distribution of Income
M. ISHAQ NADIRI AND EDWARD N. WOLFF, New York University
Short-Run Inflationary and Employment Effects of an Increase in the Minimum Wage
JACOB MINCER, Columbia University
The Economics of Wage Floors
Discussants: JACOB MINCER, Columbia University
ROBERT S. GOLDFARB, George Washington University
BRIGETTE SELLEKAERTS, Minimum Wage Study Commission
FINIS WELCH, University of California-Los Angeles

10:15 A.M. THE USES AND ABUSES OF ECONOMICS IN PUBLIC DECISION MAKING
Presiding: DICK NETZER, New York University
Papers: ELIZABETH DURBIN, New York University
Economic Theory and Political Goals: The Economics of British Democratic Socialism in the 1930's
EILEEN COLLINS, National Science Foundation
Sorting Out the Economic Arguments: Underlined Proposed Tax Incentives to Encourage Innovation

LINDA N. EDWARDS, Queens College and Graduate Center at City University of New York, AND FRANKLIN R. EDWARDS, Columbia University
Do Public Unions Have "Excessive" Power? A Test of the Wellington-Winter Hypothesis
HARRIET ZELLNER, Integral Research, Inc. and Columbia University, AND BETH T. NIEMI, Rutgers University-Newark and Integral Research, Inc.
Discrimination and Affirmative Action: The Nature of the Economic Evidence
Discussants: HENRY AARON, The Brookings Institution
YALE M. BRAUNSTEIN, Brandeis University
MARC ROSENBLUM, U.S. Equal Opportunity Commission

10:15 A.M. DOMESTIC DETERMINANTS OF FOREIGN INVESTMENT
Presiding: WILLEM H. BUITER, University of Bristol, England
Papers: RONALD I. MCKINNON, Stanford University
The Real Exchange Rate and Saving and Investment
DAVID G. HARTMAN, Harvard University
U.S. Policy and International Capital Transfers
GEORGE M. VON FURSTENBERG, International Monetary Fund
Domestic Determinants of Net Foreign Investment by the United States
Discussants: ARNOLD C. HARBERGER, University of Chicago
ALFRED STEINHERR, International Monetary Fund and Catholic University of Louvain, Belgium
RICHARD ERB, American Enterprise Institute

12:30 P.M. JOINT LUNCHEON WITH THE AEA COMMITTEE ON THE STATUS OF WOMEN IN THE ECONOMICS PROFESSION AND THE AMERICAN FINANCE ASSOCIATION
Presiding: ELIZABETH E. BAILEY, Head, CSWEP, AND FRANCO MODIGLIANI, President, American Finance Association
Speaker: ALICE M. RIVLIN, Director, Congressional Budget Office
Can Economists Help the Congress?

2:00 P.M. IMPROVING THE STATUS OF WOMEN: SOME PROVEN TECHNIQUES (Joint Session with the AEA Committee on the Status of Women in the Economics Profession)
Presiding: ELIZABETH E. BAILEY, Civil Aeronautics Board
Papers: MARY P. ROWE AND ANN F. FRIEDLAENDER, Massachusetts Institute of Technology
In Academia
NANCY M. GORDON, Senior Advisor to the Assistant to the President, The White House
In Government
RUTH SHAEFFER, The Conference Board
In Business

2:00 P.M. YOUTH EMPLOYMENT: NEW PERSPECTIVES* (Joint Session with the National Economic Association)
Presiding: BERNARD ANDERSON, University of Pennsylvania and Rockefeller Foundation
Papers: HENRY FELDER, Stanford Research Institute
Effects of Wage Subsidies on the Employment of Youth
LAURENCE MORSE, The Urban Institute
Increasing Unemployment and Changing Labor Market Expectations Among Black Male Youth
CHARLES BETSEY, The Urban Coalition, AND BRUCE DUNSON, University of Maryland
Federal Minimum Wage Laws and Employment and Earning of Minority Youth
Discussants: DAVID SWINTON, The Urban Institute
BERNARD ANDERSON, Rockefeller Foundation

2:00 P.M. ECONOMICS OF SLACK CAPACITY*
Presiding: HERBERT MOHRING, University of Minnesota
Papers: ERNST R. BERNDT, Massachusetts Institute of Technology, AND C. J. MORRISON-WHITE, University of British Columbia
Capital and Capacity Utilization in a Dynamic Model
ARTHUR DEVANY, Simon Fraser University, AND N. G. FREY, California State University-Hayward
Stochastic Equilibrium and Capacity Utilization
MURRAY F. FOSS, American Enterprise Institute
Long-Run Changes in the Work Week of Capital

WALTER Y. OI, University of Rochester
Slack Capacity: Productive or Wasteful?
Discussants: ROGER BETANCOURT, University of Maryland
JOHN C. PANZAR, Bell Telephone Laboratories

2:00 P.M. INFLATION, MONEY CREATION, AND GOVERNMENT REVENUE*
Presiding: JAMES BUCHANAN, Virginia Polytechnic Institute and State University
Papers: GEOFFREY BRENNAN AND JAMES BUCHANAN, Virginia Polytechnic Institute and State University
Revenue Implications of Money Creation under Leviathan
JEREMY SIEGEL, University of Pennsylvania
The Effects of Bank Regulation on Government Revenue from Inflation
EARL THOMPSON, University of California-Los Angeles
Who Should Control the Money Supply?
Discussants: ANDREW SCHOTTER, New York University
EDWARD TOWER, Duke University
MARTIN BAILEY, University of Maryland

2:00 P.M. TESTING IN ECONOMICS* (Sponsored by the AEA Committee on Economic Education)
Presiding: ALLEN C. KELLEY, Duke University
Papers: PHILLIP SAUNDERS, Indiana University, RENDIGS FELS, Vanderbilt University, AND ARTHUR L. WELSH, Joint Council on Economic Education
The Revised Test of Understanding College Economics (*TUCE*)
JOHN F. CHIZMAR, Illinois State University, AND JOHN C. SOPER, Northern Illinois University
Specification and Development of New Pre-College Tests: *BET* and *TEL*
Discussants: STEPHEN BUCKLES, University of Missouri-Columbia
WILLIAM B. WALSTAD, University of Missouri-St. Louis

2:00 P.M. MACRO-MICRO COMPATIBILITY: HISTORICAL AND PHILOSOPHICAL ASPECTS (Joint Session with the History of Economics Society)
Presiding: ABRAHAM HIRSCH, Brooklyn College
Papers: PAUL DAVIDSON, Rutgers University
Marshall's Economics as a Foundation for Macro Theory
E. ROY WEINTRAUB, Duke University
General Equilibrium: A Theory?
ROY J. ROTHEIM, *Challenge Magazine*
A Lesson from Walras
Discussant: WARREN SAMUELS, Michigan State University

2:00 P.M. REDISTRIBUTIVE IMPACT IN THE FEDERAL ACID RAIN ASSESSMENT PROGRAM (Joint Session with the Association for the Study of the Grants Economy)
Presiding: JANOS HORVATH, Butler University
Papers: RICHARD M. ADAMS, THOMAS D. CROCKER, AND JOHN T. TSCHIRHART, University of Wyoming
Valuing Ecosystem Functions: The Effect of Acidification
CHARLES W. HOWE, University of Colorado
International Grants via the Ecosystem: The Case of Acid Rain
MILTON KAFOGLIS, Emory University
The Redistribution of Market Share Due to Pollution Control
DOUGLAS N. ROSS, Joint Economic Committee, U.S. Congress
Approaches to Meeting Regulatory Goals: The Case of Acid Rain
Discussants: RUSSELL BOULDING, SPEA, Indiana University
WALTER P. PAGE, University of West Virginia

2:00 P.M. THE TEACHING OF ADVANCED ECONOMICS (Joint Session with Omicron Delta Epsilon)
Presiding: ALAN A. BROWN, University of Windsor
Papers: WILLIAM GILLEN, University of Windsor
The Teaching of Advanced Undergraduate and Graduate Microeconomics—A Survey
FINIS WELCH, University of California-Los Angeles
The Teaching of Advanced Labor Economics
RICHARD DUSANSKY, State University of New York-Stony Brook
The Teaching of Applied Welfare/Applied Econometrics at the Graduate Level
Discussants: JACOB MINCER, Columbia University
(Others to be announced)

2:00 P.M. ENERGY, ENVIRONMENTAL POLICY, AND ECONOMIC GROWTH
Presiding: DALE W. JORGENSON, HARVARD UNIVERSITY
Papers: EUGENE KROCH, Columbia University
Energy, Environment, and Growth
BRIAN WRIGHT, Yale University
The Cost of Aggregate Energy Conservation
WILLIAM NORDHAUS, Yale University
Energy Policy and Macroeconomic Performance
Discussants: ROBERT E. HALL, Stanford University
JOHN MORONEY, Tulane University
DAVID NISSEN, Chase Manhattan Bank

2:00 P.M. HOUSEHOLD PRODUCTION FUNCTIONS: THEORY AND ESTIMATION
Presiding: (To be announced)
Papers: F. THOMAS JUSTER, University of Michigan
The Role of Time and Wealth in the Production of Household Output
MARTHA HILL, University of Michigan
Constraints and Complementarities in Modeling Household Activities
GREG K. DOW, University of Michigan
The Dynamics of Change in the Formation of Household Preferences
FRANK P. STAFFORD, University of Michigan
The Production of Child Quality within Households
Discussants: ROBERT T. MICHAEL, University of Chicago
MICHAEL WACHTER, University of Pennsylvania
MARC NERLOVE, Northwestern University

2:00 P.M. REGULATORY FEDERALISM
Presiding: WALLACE E. OATES, University of Maryland
Papers: WERNER Z. HIRSCH, University of California-Los Angeles
Landlord-Tenant Regulations in their Local and Federal Aspects
ELI M. NOAM, Columbia University
Federalism vs. Home Rule: The Determinants of Demand for Jurisdictional Level in Urban Regulation
KENNETH T. ROSEN, University of California-Berkeley
The Effects of Zoning Laws on the U.S. Housing Market
Discussants: WILLIAM G. SHEPHERD, University of Michigan
RICHARD O. ZERBE, University of Washington
DONALD DEWEY, Columbia University

2:00 P.M. THE POLITICAL ECONOMY OF TECHNICAL CHOICE
Presiding: HERBERT GINTIS, University of Massachusetts
Papers: STEVEN COHN, University of Massachusetts
Nuclear Power: A Case Study in Official Technology
HEIDI HARTMANN, National Academy of Sciences
Technology, Patriarchy, and Housework
WILLIAM LAZONICK, Harvard University
Labor Relations and Technical Change: U.S. and British Comparisons
MICHAEL REICH, University of California-Berkeley
Political-Economic Determinants of the Division of Labor within the Firm
Discussants: SIDNEY WINTER, Yale University
PAUL DAVID, Stanford University

4:15 P.M. THE PRESIDENTIAL CANDIDATES: ECONOMIC AND FINANCIAL POLICIES (Joint Session with the American Finance Association)
Presiding: LEONARD SILK, *New York Times*
Papers: To be announced

4:15 P.M. DISTRIBUTIONAL ISSUES IN PUBLIC POLICY*
Presiding: GLENN LOURY, University of Michigan
Papers: ROBERT WILLIG, Princeton University
Social Welfare Dominance
VINCENT CRAWFORD, University of California-La Jolla
Arbitration and Conflict Resolution in Labor-Management Bargaining
Discussants: CLIFFORD DONN, University of Tennessee
ELSI KNOER, University of California-Davis

4:15 P.M. THE IMPACT OF INFLATION ON LABOR FORCE PARTICIPATION AND THE DISTRIBUTION OF HOUSEHOLD INCOME* (Joint Session with the AEA Committee on the Status of Women in the Economics Profession)
Presiding: M. LOUISE CURLEY, Scudder Stevens and Clark
Papers: CYNTHIA LLOYD AND BETH T. NIEMI, Rutgers University-Newark
Money Illusion or Price Illusion: The Effect of Inflation on Female Labor Force Participation
JUNE O'NEILL, The Urban Institute
Differential Effects of the Productivity Slowdown by Sex and Race
WILHELMINA A. LEIGH, National Urban League
Shelter Affordability: Comparisons by Race, Region, and Tenure
BARBARA BERGMANN, University of Maryland
'Economic Equity' for the Housewife
Discussants: ROBERT EISNER, Northwestern University
JOAN G. HAWORTH, Florida State University

4:15 P.M. FISCAL REFORM IN A DEMOCRACY, OR HOW LEVIATHAN MAY LOSE ITS TEETH (Joint Session with the Public Choice Society)
Presiding: ROBERT D. TOLLISON, Virginia Polytechnic Institute and State University
Papers: TERRY L. ANDERSON AND P. J. HILL, Montana State University
Fiscal Policy in the Transfer Society: A Prospective on Constitutional Reform
JUERGEN BACKHAUS, University of Constanz
The State as a Club: A Perspective for Public Finance in a Prosperous Democracy
PAUL CRAIG ROBERTS, *Wall Street Journal*
Taxation and the Demand for Leviathan
RICHARD E. WAGNER, Auburn University
Taxes, Prices, and the Institutional Requisites for Economic Calculation
Discussants: DAVID E. R. GAY, University of Chicago Law School
KAREN I. VAUGHN, George Mason University

4:15 P.M. THE ANTI-SMOKING CAMPAIGN (Joint Session with the Health Economics Research Organization)
Presiding: DONALD YETT, University of Southern California
Papers: JEFFREY E. HARRIS, Massachusetts Institute of Technology
Quitting Cigarette Smoking
DOUGLAS C. COATE, Rutgers University, AND EUGENE M. LEWIT, College of Medicine and Dentistry of New Jersey-New Jersey Medical School
The Potential for Reducing Cigarette Smoking Through the Use of Excise Taxes
KENNETH E. WARNER, University of Michigan
Benefits and Costs of the Anti-Smoking Campaign: Framework for Assessment
Discussants: DONALD SANT, Federal Trade Commission
MARVIN KRISTEIN, State University of New York-Stony Brook
STUART SCHWEITZER, University of California-Los Angeles

4:15 P.M. INTERNATIONAL DIMENSIONS OF ECONOMIC EDUCATION
Presiding: EDWIN S. MILLS, Princeton University
Papers: WYN OWEN, University of Colorado
Components of the International Dimension of Economic Education
GUSTAV RANIS, Yale University, EDWARD SCHUH, University of Minnesota, JAMES WORLEY, Vanderbilt University, AND PHILIP CATEURA, University of Colorado
Panel Discussion: Developing the International Dimension

4:15 P.M. FINANCIAL DEVELOPMENT AND STABILIZATION POLICIES IN *LDC*s
Presiding: RONALD I. MCKINNON, Stanford University
Papers: MAXWELL FRY, University of Hawaii
Financial Development Models: Theories, Tests, and Extensions
ANDREW CROCKETT, International Monetary Fund
Implementation of Financial Stabilization in Developing Countries: Selected Cases
PEKKA AHTIALA, University of Tampere, Finland, AND JEAN-MARIE BLIN, Northwestern University
Risk, Resource Allocation, and the Social Cost of Inflation
FREDERICK BERGER, Crocker Bank of San Francisco
Financial Immaturity and Monetary Control in an Open Economy
Discussants: DONALD MATHIESON, International Monetary Fund
RONALD I. MCKINNON, Stanford University

4:15 P.M. CURRENT RESEARCH IN INTERNATIONAL BUSINESS CYCLES
Presiding: GEOFFREY H. MOORE, Rutgers University-Newark
Papers: VICTOR ZARNOWITZ, University of Chicago, AND GEOFFREY H. MOORE, Rutgers University-Newark
Trigger Formulas for Recognizing Recessions and Recoveries
JOHN P. CULLITY, Rutgers University-Newark
Cost-Price Relationships in Business Cycles in Industrial Countries
CHRISTOPHER SIMS, University of Minnesota
Dynamic Business Cycle Models for Several Countries
NEIL SHEFLIN, Rutgers University
Leading Indicators of Inflation
Discussants: (To be announced)

8:00 P.M. RICHARD T. ELY LECTURE*
Presiding: WILLIAM J. BAUMOL, Princeton University and New York University
Speaker: LORD LIONEL ROBBINS, London School of Economics
(Title to be announced)

Saturday, September 6, 1980

8:00 A.M. THE COST OF SMOG (Joint Session with the Association of Environmental and Resource Economists)
Presiding: CHARLES HOWE, University of Colorado
Papers: DAVID BROOKSHIRE, RALPH D'ARGE, WILLIAM SCHULZE, University of Wyoming, AND MARK THAYER, University of Missouri
Benefits of Reducing Los Angeles Smog
ANNE BRESNOCK, University of Colorado and San Diego State University
Housing Prices, Income, and Environmental Quality in Denver
EDNA LOEHMAN, RON PROMBOIN, KATHLEEN CHAIKIN, AND DAVID BOLDT, Stanford Research Institute
Experiments in Measurement of Air Pollution Benefits in San Francisco
Discussants: RONALD CUMMINGS, University of New Mexico
A. MYRICK FREEMAN III, Bowdoin College

8:00 A.M. INFLATION AND THE TAX SYSTEM*
Presiding: MARTIN FELDSTEIN, Harvard University and National Bureau of Economic Research
Papers: ALAN AUERBACH, Harvard University
Inflation and Firm Behavior
LAWRENCE SUMMERS, Massachusetts Institute of Technology
Inflation and the Valuation of Physical Capital
MARTIN FELDSTEIN, Harvard University and National Bureau of Economic Research
Inflation and Private Pensions
Discussant: PATRIC HENDERSHOTT, Purdue University

8:00 A.M. A CRITICAL LOOK AT RATIONAL EXPECTATIONS IN PRACTICE* (Joint Session with the Society of Economic Dynamics and Control)
Presiding: JOHN B. TAYLOR, Columbia University
Papers: GREGORY C. CHOW, Princeton University
Estimation and Control of Rational Expectations Models
MAURICE OBSTFELD, Columbia University
Rational Expectations and Financial Policy in Open Economies
ROBERT J. SHILLER, University of Pennsylvania
Volatility Measures and Rational Expectations Models
Discussants: ALBERT ANDO, University of Pennsylvania
DALE HENDERSON, Board of Governors, Federal Reserve System
ROBERT LUCAS, University of Chicago

8:00 A.M. TOPICS IN PUBLIC CHOICE (Joint Session with the Public Choice Society)
Presiding: RICHARD B. MCKENZIE, Clemson University
Papers: RICHARD CEBULA, Emory University, AND JAMES V. KOCH, Emory University
An Analysis of Selected Dimensions of Public Income Redistribution Policies in the United States During the 1960's and the 1970's
ROGER FAITH AND JAMES BUCHANAN, Virginia Polytechnic Institute and State University
Majority Voting on Yes-No Issues: Analogues and Contrasts with Voting on *N*-Alternatives

DELORES MARTIN, University of Nebraska-Lincoln
Weather Forecasting Services: Public and Private Choice
Discussants: MORGAN REYNOLDS, Texas A&M University
WILLIAM COBB, West Virginia College of Graduate Studies
HUGH MACAULAY, Clemson University

8:00 A.M. APPLICATIONS OF LONGITUDINAL PANEL DATA
Presiding: JAMES N. MORGAN, University of Michigan
Papers: HERBERT PARNES, Ohio State University
An Overview of Results from the NLS
SAUL HOFFMAN, University of Delaware, AND GREG DUNCAN, University of Michigan
The Dynamics of Wage Rate Change
RICHARD COE AND MARTHA HILL, University of Michigan
The Dynamics of Welfare Program Participation and Poverty Status
LINDA DOTCHER AND MARY CORCORAN, University of Michigan
The Role of Family Background in the Generation of Economic Status
Discussants: PAUL TAUBMAN, University of Pennsylvania
ZVI GRILICHES, Harvard University
ROBERT H. HAVEMAN, University of Wisconsin-Madison

10:15 A.M. ARE MARKETS EFFICIENT? TESTS OF ALTERNATIVE HYPOTHESES* (Joint Session with the American Finance Association)
Presiding: IRWIN FRIEND, University of Pennsylvania
Papers: RICHARD A. COHN, University of Illinois at Chicago Circle, AND DONALD R. LESSARD, Massachusetts Institute of Technology
The Effect of Inflation on Stock Prices: International Evidence
ROBERT J. SHILLER, University of Pennsylvania
The Relationship between Financial Asset Prices and Subsequent Earnings
Discussants: JOHN LINTNER, Harvard University
JOHN B. LONG, University of Rochester
JAMES R. F. GUY, Lehman Bros. Kuhn Loeb Investments

10:15 A.M. RESOURCE ALLOCATION IN THE LOCAL PUBLIC SECTOR*
Presiding: HELEN F. LADD, Harvard University
Papers: DENNIS EPPLE, Carnegie-Mellon University, AND ALLAN ZELENITZ, Tulane University
Equilibria in Some Local Public Finance Models
WALLACE E. OATES, University of Maryland
On Local Finance and the Tiebout Model
JOHN YINGER, Harvard University
Capitalization and the Median Voter
Discussants: PETER MIESZKOWSKI, University of Houston
JANET PACK, University of Pennsylvania
GEORGE ZODROW, Rice University

10:15 A.M. MEASURING BENEFITS FROM AIR POLLUTION CONTROL: WHERE DO WE STAND?*
Presiding: ALLEN V. KNEESE, Resources for the Future, Inc.
Papers: WILLIAM SCHULTZE, University of Wyoming
Benefits from Reduced Mortality
MAUREEN CROPPER, University of Southern California
Benefits from Reduced Morbidity
SHERWIN ROSEN, University of Chicago
Valuing Risk
Discussant: A. MYRICK FREEMAN III, Bowdoin College

10:15 A.M. MONETARY THEORY: IMPLICATIONS FOR MONETARY POLICY*
Presiding: ANDREW BRIMMER, Brimmer and Company, Inc.
Papers: STEPHEN H. AXILROD AND DAVID E. LINDSEY, Federal Reserve System
Implementation of Monetary Policy: Analytical Foundations of the New Approach
JACOB A. FRENKEL AND MICHAEL MUSSA, University of Chicago
Monetary Policy in an Open Economy
ANDREW BRIMMER, Brimmer and Company, Inc., AND ALLEN SINAI, Data Resources, Inc.
Rational Expectations and the Conduct of Monetary Policy
Discussants: RAYMOND E. LOMBRA, Pennsylvania State University
RONALD I. MCKINNON, Stanford University

10:15 A.M. Sustainability Analysis*
Presiding: Edward Zajac, Bell Laboratories
Papers: John C. Panzar, Bell Laboratories, and Robert Willig, Princeton University
Economies of Scope
Kenneth Baseman, U.S. Department of Justice
Sustainability and the Entry Process
Janusz Ordover and Andrew Schotter, New York University
On the Political Sustainability of Taxes
Discussants: David Teece, Stanford University, Graduate School of Business
Gunter Knieps, University of Bonn
Sam Peltzman, University of Chicago, Graduate School of Business

10:15 A.M. Issues in Government Economic Regulation of Transportation in the 1980's* (Joint Session with the Transportation and Public Utilities Group of the AEA)
Presiding: Donald V. Harper, University of Minnesota
Papers: John C. Spychalski, Pennsylvania State University
Antitrust Standards and Railway Freight Pricing: New Round in an Old Debate
Benjamin J. Allen, Iowa State University
The Nature, Effectiveness, and Importance of Motor Common Carrier Service Obligations
John Guandolo, MacDonald and McInerny, P.C.
The Role of the Interstate Commerce Commission
Discussant: Ernest W. Williams, Jr., Columbia University

10:15 A.M. Technical Change
Presiding: Robert Crouch, University of California-Santa Barbara
Papers: Dale W. Jorgenson, Harvard University, and Barbara M. Fraumeni, Boston College
Technical Change and Energy Prices
Edwin Mansfield, University of Pennsylvania
Imitation Costs, Patents, and Market Structure
Murray Brown and Richard Greenberg, State University of New York-Buffalo
Technical Change and Aggregation Bias
Robert Crouch, University of California-Santa Barbara
Learning, Technical Change, and the Aggregate Production Function
Discussants: Robert M. Solow, Massachusetts Institute of Technology
M. Ishaq Nadiri, New York University
Nestor E. Terleckyj, National Planning Association

10:15 A.M. Economics of Information
Presiding: Helene Ebenfield, National Science Foundation
Papers: Beth Elaine Allen, University of Pennsylvania
Differential Information in Pure Exchange Economies
Yale M. Braunstein and Lawrence B. Pulley, Brandeis University
Market Structure in Information Industries: An Empirical Study
Joseph E. Stiglitz, Princeton University
Information and Economic Incentives

10:15 A.M. Empirical Public Choice (Joint Session with the Public Choice Society)
Presiding: Robert D. Tollison, Virginia Polytechnic Institute and State University
Papers: Michael Maloney, Clemson University, and Robert McCormick, University of Rochester
Property Rights and Cartels: Rent Distribution
Cotton M. Lindsey, University of California-Los Angeles and Emory University, and Stuart Feigenbaum, University of California-Los Angeles
Buying Votes with the Public Purse
Robert Mackay and Carolyn Weaver, Virginia Polytechnic Institute and State University
Bureau Consolidation and Growth of Government: An Empirical Investigation
Discussants: Ronald Batchelder, Tulane University
Barry Baysinger, Texas A&M University
Gary Libecap, Texas A&M University

10:15 A.M. Contemporary Economics as a Chapter in the History of Economic Thought (Joint Session with the History of Economics Society)
Presiding: Royall Brandis, University of Illinois-Urbana
Papers: Leonard Silk, *New York Times*
(Title to be announced)

DANIEL FUSFELD, University of Michigan
(Title to be announced)
INGRID H. RIMA, Temple University
(Title to be announced)
Discussant: KENNETH E. BOULDING, University of Colorado

10:15 A.M. POLICY ISSUES IN FINANCING SMALL AND MINORITY BUSINESSES (Joint Session with the National Economic Association)
Presiding: ALFRED EDWARDS, University of Michigan
Papers: WILLIAM DUNKELBERG, Purdue University
Small Business Indicators of Credit Availability
WILLIAM BRADFORD, Stanford University
301(d) SBICs and Minority Business Finance
TIMOTHY BATES, University of Vermont
Effectiveness of the SBA in Financing Minority Businesses
STEVEN E. LEVY AND ALFRED E. OSBORNE, University of California-Los Angeles
Securities Regulation and Small Business Finance
Discussants: WILLIAM HUNTER, University of Georgia
DAVID BROPHY, University of Michigan
LUCY RUBEN, Duke University

10:15 A.M. COMPARATIVE DIMENSIONS OF GRANTS ECONOMICS (Joint Session with the Association for the Study of the Grants Economy and the Association for Comparative Economic Studies)
Presiding: EGON NEUBERGER, State University of New York-Stony Brook
Papers: SUSAN ROSE-ACKERMAN, Yale University
Charity Market: Paying Customers and Quality Control
ESTELLE JAMES, State University of New York-Stony Brook
The Non-Profit Sector in Less Developed Economies
TIBOR PALANKEI, Karl Marx University of Economics, Budapest
Measuring the Size and Leverage of Subsidies in Centrally Planned Economies
Discussants: HENRY HANSMANN, University of Pennsylvania Law School
JOHN MICHAEL MONTIAS, Yale University
ALAN A. BROWN, Windsor University

10:15 A.M. THE EFFECT OF SOCIAL SECURITY ON SAVING AND BEQUESTS
Presiding: MARTIN FELDSTEIN, Harvard University
Papers: PAUL MENCHIK, Michigan State University, AND MARTIN DAVID, University of Wisconsin
The Effect of Social Security on Bequests
ALAN S. BLINDER AND ROGER GORDON, Princeton University
Social Security and Asset Accumulation of Older Men
MICHAEL BOSKIN, Stanford University
Social Security and Saving
DEAN LEIMER AND SELIG LESNOY, U.S. Social Security Administration
Social Security and Private Saving
Discussants: LAWRENCE KOTLIKOFF, University of California-Los Angeles
ANTHONY PELLECHIO, University of Rochester
HENRY AARON, The Brookings Institution

12:30 P.M. LUNCHEON IN HONOR OF THE 1979 NOBEL LAUREATES, W. ARTHUR LEWIS AND THEODORE SCHULTZE
Presiding: MOSES ABRAMOVITZ, Stanford University
Speakers: WILLIAM J. BAUMOL, Princeton University and New York University
D. GALE JOHNSON, University of Chicago

2:00 P.M. INCOMES POLICY: WHAT DID WE LEARN IN THE 1970'S?*
Presiding: ALBERT REES, Alfred P. Sloan Foundation
Papers: D. QUINN MILLS, Harvard University
U.S. Incomes Policy in the 1970's: Assumptions, Objectives and Results
JON FRYE AND ROBERT J. GORDON, Northwestern University
An Econometric Retrospective on Postwar U.S. Incomes Policies
LAURENCE S. SEIDMAN, University of Pennsylvania
Tradeoffs in the Design of an Incomes Policy
Discussants: ALAN S. BLINDER, Princeton University
ROBERT W. CRANDALL, The Brookings Institution

2:00 P.M. ECONOMICS OF AFFIRMATIVE ACTION*
Presiding: PERRY SHAPIRO, University of California-Santa Barbara
Papers: GLENN LOURY, University of Michigan
Arguments for Affirmative Action
FINIS WELCH, University of California-Los Angeles
(Title to be announced)
Discussants: FINIS WELCH, University of California-Los Angeles
GLENN LOURY, University of Michigan

2:00 P.M. BEHAVIORAL-MICRO APPROACHES TO INFLATION*
Presiding: HARVEY LEIBENSTEIN, Harvard University
Papers: SHLOMO MAITAL, Princeton University
Inflation Expectations and the Changing Nature of Borrowing and Credit
JAMES W. DEAN, Columbia University
Some Gaps in Conventional Inflation Theory
HARVEY LEIBENSTEIN, Harvard University
Aspects of the Micro-Economics of Inflation: An X-Efficiency Theory Approach
Discussants: PAUL DAVIDSON, Rutgers University-New Brunswick
PATRICK CLAWSON, Seton Hall University

2:00 P.M. FREIGHT TRANSPORT REGULATION AND MARKET IMPERFECTIONS*
Presiding: DARIUS W. GASKINS, Interstate Commerce Commission
Papers: RUSSELL C. CHERRY, Arthur D. Little, Inc.
Market Structure and Concentration in the Regulated Trucking Industry
ANN F. FRIEDLAENDER, Massachusetts Institute of Technology
Price Distortions and Second Best Investment Rules in the Transport Industries
RICHARD C. LEVIN, Yale University
Railroad Regulation and Workable Competition
Discussants: RONALD BRAUETIGAM, Northwestern University
ANNETTE LAMOND, Charles River Associates
CLIFFORD WINSTON, Massachusetts Institute of Technology

2:00 P.M. PANEL DISCUSSION IN HONOR OF PAUL SAMUELSON: ECONOMIC THEORY AND POLICY
Presiding: GEORGE R. FEIWEL, University of Tennessee
Panelists: KENNETH J. ARROW, Stanford University
WILLIAM J. BAUMOL, Princeton University and New York University
LAWRENCE KLEIN, University of Pennsylvania
ROBERT M. SOLOW, Massachusetts Institute of Technology

2:00 P.M. INTERNATIONAL COMPARISONS OF INCOME CONSUMPTION AND WELFARE (Joint Session with the Association for Comparative Economic Studies)
Presiding: BELA BALASSA, Johns Hopkins University and World Bank
Papers: IRVING B. KRAVIS, ALAN HESTON, AND ROBERT SUMMERS, University of Pennsylvania
Changes in the World Income Distribution
ROBIN MARRIS, University of Maryland, AND HENRI THEIL, University of Chicago
International Comparisons of Economic Welfare
LAURITS R. CHRISTENSEN, DIANNE CUMMINGS, University of Wisconsin-Madison, AND DALE W. JORGENSON, Harvard University
An International Comparison of Levels of Total Factor Productivity
Discussants: ABRAM BERGSON, Harvard University
W. ERWIN DIEWERT, University of British Columbia

2:00 P.M. RECENT ECONOMISTS: THEIR WORK AND THEIR LEGACY (Joint Session with the History of Economics Society)
Presiding: PATRICIA BOWERS, Brooklyn College
Papers: EUGENE ROTWEIN, Queens College
Jacob Viner and the Chicago Tradition
LAWRENCE MOSS, Babson College, AND KAREN I. VAUGHN, George Mason University
Ludwig von Mises: The Austrian Tradition
ANN SCHWIER, Southern Illinois University at Edwardsville, AND JEROME F. SCHWIER, St. Louis
Mr. Phillips and His Curve
Discussants: (To be announced)

2:00 P.M. RISK, LIABILITY, AND NEW TECHNOLOGY IN PUBLIC UTILITIES (Joint Session with the Transportation and Public Utilities Group of the AEA)
Presiding: JAMES SUELFLOW, Indiana University
Papers: JOHN D. LONG, Indiana University
The Limitations of Nuclear Insurance as a Risk Reduction Device in Electricity Production
WILLIAM C. WOOD, Vanderbilt University
Responses of Electric Utilities to Changes in Liability for Nuclear Accidents
Discussants: TERRY A. FERRAR, Edison Electric Institute
RODNEY E. STEVENSON, University of Wisconsin-Madison

2:00 P.M. ACHIEVING ECONOMIC EFFICIENCY AND EQUITY IN INTERNATIONAL AGRICULTURAL TRADE AND DEVELOPMENT (Joint Session with the American Agricultural Economics Association)
Presiding: KENNETH NOBE, Colorado State University
Papers: RAY GOLDBERG, Harvard University
The Role of the Multinational Corporation
JIMMYE HILLMAN, University of Arizona
The Role of Export Cropping in Less Developed Countries
ALAIN DE JANVRY, University of California-Berkeley
The Role of Land Reform
Discussants: TERRY OZAWA, Colorado State University
MALCOLM BALE, World Bank
DANIEL BROMLEY, University of Wisconsin

2:00 P.M. TOPICS IN ECONOMICS OF MEDIA (Joint Session with Omicron Delta Epsilon)
Presiding: MICHAEL SZENBERG, Long Island University
Papers: STEVEN WILDMAN, University of California-Los Angeles
The Economic Consequences of the Informational Characteristics of Mass Media
JOE GRUNDFEST, Stanford University
Regulation in the Face of Uncertainty
DAVID WATERMAN, University of Southern California and Occidental College
Public Goods and the Structure of the Motion Picture Industry
JAMES DERTOUZOS, Rand Corporation
Scale Economies, Newspaper Chains, and Government Policy
Discussants: STEVE MALIN, City University of New York and The Conference Board
STEPHEN KAGANN, St. Johns University
KENNETH LEESON, New York University and National Telecommunications and Information Administration

2:00 P.M. APPLICATIONS OF REGIONAL ECONOMETRIC MODELS
Presiding: WILLIAM J. LAWRENCE, Pace University
Papers: BENJAMIN CHINITZ, Abt Associates, AND ROGER BOLTON, Williams College
Multiregional Models in Policy Planning: A Survey
DAVE KRESGE, Harvard-M.I.T. Joint Center for Urban Studies
(Title to be announced)
DAVID BETSON, SHELDON DANZIGER, AND ROBERT H. HAVEMAN, University of Wisconsin-Madison
Estimating the Regional and Distributional Impacts of Wage Subsidy Policies
LEONARD FORMAN, *New York Times*, AND WILLIAM J. LAWRENCE, Pace University
Linking Regional Econometric Models and Corporate Planning Models
Discussants: (To be announced)

2:00 P.M. DISEQUILIBRIUM ECONOMICS AND THE EVOLUTIONARY VISION
Presiding: FRITZ MACHLUP, Princeton University and New York University
Papers: KENNETH E. BOULDING, University of Colorado
From Entropy to Autopoiesis: Contributions to Economics from Other Disciplines
RICHARD H. DAY, University of Southern California
Crises and Chaos in Economic Development: A General Theory of the Dynamic Economy
SIDNEY G. WINTER, Yale University
Persistent Inefficiencies in Economic Evolutionary Processes
BRADFORD CORNELL AND RICHARD ROLL, University of California-Los Angeles
Evolutionary Stable Strategies in Markets and Economic Organizations
Discussants: (To be announced)

4:15 P.M. MODELS OF LABOR MARKET EQUILIBRIUM*
Presiding: DALE T. MORTENSEN, Northwestern University
Papers: SANFORD GROSSMAN, University of Pennsylvania, AND OLIVER HART, Churchill College, Cambridge University
Contracting Subject to Asymmetric Information
BENGT HOLMSTROM, Northwestern University
Long-Term Contracts and Labor Market Equilibrium
KENNETH BURDETT, University of Wisconsin-Madison, AND DALE T. MORTENSEN, Northwestern University
Immobility and Labor Market Equilibrium
Discussants: (To be announced)

4:15 P.M. THE MONETARIST CONTRIBUTION TO MACROECONOMICS*
Presiding: JEROME L. STEIN, Brown University
Papers: BENNETT MCCALLUM, University of Virginia
Monetarism and Rational Expectations
JEROME L. STEIN, Brown University
Monetarist, Keynesian, and the New Classical Economics
JOHN B. TAYLOR, Columbia University
High Inflation and Monetary Stability
Discussants: ALBERT ANDO, University of Pennsylvania
KARL BRUNNER, University of Rochester
CARL F. CHRIST, Johns Hopkins University

4:15 P.M. THE PRODUCTIVITY GROWTH SLOWDOWN*
Presiding: ERNST R. BERNDT, Massachusetts Institute of Technology
Papers: M. ISHAQ NADIRI, New York University
Technical Change, Returns to Scale, and the Growth Slowdown
G. CHRISTAINSEN AND ROBERT H. HAVEMAN, University of Wisconsin-Madison
Regulation and Productivity Growth: The Nature and Extent of the Interaction
MARTIN N. BAILY, The Brookings Institution
What Do We Know About the Productivity Growth Slowdown?
Discussants: ERNST R. BERNDT, Massachusetts Institute of Technology
RICHARD C. LEVIN, Yale University
DALE W. JORGENSON, Harvard University

4:15 P.M. LAW AND ECONOMICS IN AN ENVIRONMENT WITH IMPERFECT INFORMATION*
Presiding: ANDREW SCHOTTER, New York University
Papers: JANUSZ ORDOVER, New York University, AND ANDREW WEISS, Bell Laboratories
An Economic Basis of Unconscionability
RICHARD A. POSNER, University of Chicago Law School
The Economics of Privacy
STEVEN SALOP AND HOWARD BEALES, Federal Trade Commission
Consumer Information Remedies
STEVEN SHAVELL, Harvard University
Efficiency and Equity in Legal Rule Making: Should Equity Matter Given Optimal Income Tax?
Discussant: LEWIS KORNHAUSER, New York University Law School

4:15 P.M. SIZE AND POLICY IMPLICATIONS OF THE "UNDERGROUND ECONOMY"
Presiding: PAUL B. MANCHESTER, Joint Economic Committee, U.S Congress
Papers: VITO TANZI, International Monetary Fund
Impact of the Underground Economy on Tax Evasion
BERDJ KENADJIAN, U.S. Internal Revenue Service
The Direct Approach to Measuring the Underground Economy: IRS Estimates of Unreported Income
EDGAR FEIGE, University of Wisconsin-Madison
The Theory and Measurement of the Unobserved Sector of the U.S. Economy: Causes, Consequences, and Implications
J. S. CRAMER, University of Amsterdam
Stocks and Flows in the Irregular Economy
Discussants: RICHARD PORTER, U.S. Federal Reserve System
ROBERT EISNER, Northwestern University
JAMES HENRY, McKinsey and Company
BARRY MOLEFSKY, Congressional Research Service, U.S. Library of Congress

4:15 P.M. ECONOMIC GROWTH AND THE QUALITY OF LIFE
Presiding: RONALD G. RIDKER, Resources for the Future
Papers: ROLAND N. MCKEAN, University of Virginia
The Economics of Negative Growth
WALTON T. WILFORD AND DAVID A. LARSON, University of New Orleans
On Measuring the Quality of Life
WALTER W. HAINES, New York University
Keynes and His Grandchildren: A 50th Anniversary Reflection on Growth and Life
Discussants: EVSEY D. DOMAR, Massachusetts Institute of Technology
MARK A. LUTZ, University of Maine
A. MYRICK FREEMAN III, Bowdoin College

4:15 P.M. ECONOMIC ANALYSIS WITHIN A CAPITALIST FRAMEWORK
Presiding: HYMAN P. MINSKY, Washington University, St. Louis
Papers: JAN KREGEL, Rutgers University, and Rijksuniversiteit te Groningen, The Netherlands
Expectations and Rationality in a Capitalist Framework
EDWARD NELL, New School for Social Research
Pricing and Investment in Competitive and Corporate Markets
JACOB COHEN, University of Pittsburgh
A Flow Constrained Approach to Monetary Theory
Discussants: NEIL WALLACE, University of Minnesota
E. ROY WEINTRAUB, Duke University
OLIVIER BLANCHARD, Harvard University

4:15 P.M. THE NEW ECONOMICS OF LABOR UNIONS
Presiding: JOSEPH D. REID, JR., Virginia Polytechnic Institute and State University
Papers: ROGER L. FAITH AND JOSEPH D. REID, JR., Virginia Polytechnic Institute and State University
The Labor Union as its Members' Agent
HENRY S. FARBER, Massachusetts Institute of Technology
An Analysis of Hicks' Theory of Industrial Disputes
JAMES B. KAU AND PAUL H. RUBIN, University of Georgia
The Political Influence of Unions on Congressional Voting
Discussants: H. GREGG LEWIS, Duke University
MELVIN W. REDER, University of Chicago

4:15 P.M. EXPERIMENTAL METHODS APPLIED TO EMPLOYMENT AND TRAINING: SUPPORTED WORK
Presiding: ROBIN G. HOLLISTER, Swarthmore College and Mathematica Policy Research, Inc.
Papers: REBECCA MAYNARD, Mathematica Policy Research, Inc.
Youth: A Test of Direct Employment
KATHY DICKINSON AND IRVING PILIAVIN, Institute for Research on Poverty, University of Wisconsin-Madison
Ex-Addicts and Ex-Offenders: Employment and Recidivism
STANLEY MASTERS, Institute for Research on Poverty, University of Wisconsin-Madison
Women from AFDC: Work and Welfare
PETER KEMPER, DAVID LONG, AND CRAIG THORNTON, Mathematica Policy Research, Inc.
A Benefit-Cost Analysis of the Supported Work Experiment
Discussants: ERNST W. STROMSDORFER, Abt Associates
MICHAEL WISEMAN, University of California-Berkeley

4:15 P.M. DEMANDS FOR HEALTH CARE SERVICES (Joint Session with the Health Economics Research Organization)
Presiding: LOUIS ROSSITER, National Center for Health Services, U.S. Department of Health, Education, and Welfare
Papers: WILLARD G. MANNING, JOSEPH P. NEWHOUSE, C. N. MORRIS et al., Rand Corporation
Experimental Results on the Demand for Medical Care Services: Findings from the Health Insurance Study
STUART GUTERMAN, KATHRYN M. LANGWELL, AND ROBERT T. DEANE, Applied Management Sciences, Inc.
Utilization of Dental Services at Zero Price and Full Price: A Comparative Examination of Two Groups
DEBORAH A. FREUND, University of North Carolina
Black-White Differentials and the Demand for Medical Care: How Close Are We to Parity?
Discussants: HERBERT E. KLARMAN, New York University
JOSEPH LIPSCOMB, Duke University
BERNADETTE CHACHERE, Health Care Financing Administration, U.S. Department of Health, Education, and Welfare

4:15 P.M. PRICING AND COMPETITION IN PUBLIC UTILITIES (Joint Session with the Transportation and Public Utilities Group of the AEA)
Presiding: JOHN C. SPYCHALSKI, Pennsylvania State University
Papers: MANLEY R. IRWIN, University of New Hampshire
Price and Non-Price Competition in Communications: The Next Ten Years
JACKALYNE P. SMITH, California Public Utilities Commission
Impact of New Departures in Electricity Pricing
Discussants: EDYTHE S. MILLER, Colorado Public Utilities Commission
WILLIAM VICKREY, Columbia University

8:00 P.M. PRESIDENTIAL ADDRESS
Presiding: KENNETH J. ARROW, Stanford University
Speaker: MOSES ABRAMOVITZ, Stanford University

9:30 P.M. BUSINESS MEETING

Sunday, September 7, 1980

8:00 A.M. APPLICATIONS OF PRODUCTION THEORY AND THEORY-CHOICE ANALYSIS
Presiding: JAMES CICARELLI, State University of New York-Oswego
Papers: BRUCE J. CALDWELL, University of North Carolina-Greensboro
The Analysis of Theory-Choice in Economics: Some Possible Alternatives
JOHN B. LOOMIS, U.S. Fish and Wildlife Service
Increasing Commodity and Amenity Production from Government Owned Land: Lessons from Trade Theory
WILLIAM L. MARR, Wilfrid Laurier University
The Wheat Economy in Reverse: Ontario's Wheat Production, 1887–1917
LAWRENCE SPIZMAN AND JAMES CICARELLI, State University of New York-Oswego
The Production of Economic Knowledge
Discussants: JACK EDWARDS, U.S. Bureau of Land Management
VINCENT TARASCIO, University of North Carolina-Chapel Hill
STEPHEN DECANIO, University of California-Santa Barbara
JOHN J. SIEGFRIED, Vanderbilt University

8:00 A.M. ESTIMATING THE DEMAND FOR GOODS OF VARYING QUALITY
Presiding: ROBERT MENDELSOHN, University of Washington
Papers: JACK TRIPLETT, U.S. Bureau of Labor Statistics
Nonlinear Budget Constraints in Characteristics Space
GARDNER BROWN AND ROBERT MENDELSOHN, University of Washington
The Demand for Hunting of Varying Qualities
JOHN QUIGLEY, University of California-Berkeley
Consumer Demand for the Attributes of Housing Quality
JOHN KAIN, Harvard University
(Title to be announced)
Discussants: A. MYRICK FREEMAN III, Bowdoin College
(Others to be announced)

10:15 A.M. THE ECONOMICS OF RECESSIONS*
Presiding: ROBERT E. HALL, Stanford University
Papers: OLIVIER BLANCHARD, Harvard University
Is There Such a Thing as an Anticipated Recession?
BEN BERNANKE, Stanford University
Liquidity Constraints, Bankruptcy, and Recession
RAY FAIR, Yale University
Estimated Effects of the October 1979 Changes in Monetary Policy on the 1980 Economy
Discussants: LAURENCE WEISS, Yale University
PAUL EVANS, Stanford University

10:15 A.M. NEW DEVELOPMENTS IN THE DESIGN OF MONETARY POLICY (Joint Session with the American Finance Association)
Presiding: CARL F. CHRIST, Johns Hopkins University
Papers: PETER TINSLEY, BONNIE GARRETT, ALFRED NORMAN, P. A. V. B. SWAMY, AND PETER VON ZUR MUHLEN, Board of Governors of the Federal Reserve System
The Impact of Uncertainty on the Feasibility of the Humphrey-Hawkins Objectives

David Pierce, William Cleveland, Agustin Maravall, and Darrell Parke, Board of Governors of the Federal Reserve System
Uncertainty in the Monetary Aggregates: Sources, Measurement, and Policy Effects
William Barnett, Paul Spindt, and Edward Offenbacher, Board of Governors of the Federal Reserve System
New Concepts of Aggregated Money and their Expected Information Content
George Moore, William Conrad, and Garry Schinasi, Board of Governors of the Federal Reserve System
A Non-Neutral View of Neutrality: The Cost of Capital and Monetary Policy
Discussants: Robert M. Solow, Massachusetts Institute of Technology
Louis J. Maccini, Johns Hopkins University
Christopher Sims, University of Minnesota
W. Erwin Diewert, University of British Columbia

10:15 A.M. ISSUES IN URBAN TRANSPORTATION IN THE 1980's (Joint Session with the Transportation and Public Utilities Group of the AEA)
Presiding: Frederick J. Stephenson, University of Georgia
Papers: Ray A. Mundy, University of Tennessee
Public Mass Transportation Management Trends in the 1980's
Robert C. Lieb, Northeastern University
Urban Transportation: Labor in the 1980's
Dennis R. McDermott, Virginia Commonwealth University
Urban Goods Movement: State of the Art and Future Perspectives
Discussant: George M. Smerk, Indiana University

10:15 A.M. WORLD INFLATION
Presiding: Myron J. Frankman, McGill University
Papers: Jaleel Ahmad, Concordia University
Floating Exchange Rates and World Inflation
Russell Boyer, Carnegie-Mellon University and Geoffrey Kingston, University of Western Ontario
Fiscal Deficits, Monetary Growth and Inflation Rates: U.S. and Canadian Postwar Experience
Douglas Purvis, Queen's University
International Aspects of Inflation
M. Sakbani, United Nations Conference on Trade and Development
Trends and Causes of Inflation in Developing Countries
Discussants: James Riedel, Johns Hopkins University
Robert M. Dunn, Jr., George Washington University
Walter Enders, Iowa State University
Craig MacPhee, University of Nebraska

10:15 A.M. MANAGING MUNICIPAL SERVICES IN THE ABSENCE OF COMMUNITY GROWTH
Presiding: Richard E. Schuler, Cornell University
Papers: Walter Lynn and Richard Schuler, Cornell University
Shrinking Capital Intensive Public Facilities: Theoretical and Institutional Issues
George Tolley and Larry Huckins, University of Chicago
Explaining Municipal Capital Expenditures under Economic Change
Roy Bahl and Larry Schroeder, Syracuse University
Adjusting State and Local Government Budgets During Periods of Slow Economic Growth
Wilbur Thompson, Wayne State University
Balanced Contraction: A Synthesis
Discussants: John Quigley, University of California-Berkeley
Benjamin Chinitz, Abt Associates and State University of New York-Binghamton
George Peterson, The Urban Institute

Editor's Note

*Papers from sessions marked with an asterisk will be published in the *Papers and Proceedings* issue of the *Review*, May 1981.

ANNOUNCEMENTS

The ninety-third annual meeting of the American Economic Association will be held in Denver, Colorado, September 5-7, 1980. The Professional Placement Service will be held December 28-30, 1980 at the Dallas Hilton, Dallas, Texas. Operating hours will be December 28, 10:00 A.M.–5:00 P.M., December 29, 9:00 A.M.–5:00 P.M., December 30, 9:00 A.M.–2:00 P.M.

Requests for placement forms should be made to Ms. Janet Aschenbrenner, National Registry for Economists, Illinois Job Service, 40 West Adams Street,Chicago, IL 60603 between September 1 and November 15. *Note*: No forms will be mailed prior to September; however, they will be available at the American Economic Association booth at the annual meeting. Completed forms *must* be returned by December 1. You do not have to attend the meeting to have your listing on file. There is no registration fee.

Resolutions for Consideration at the Annual Business Meeting

The Executive Committee at its meeting on March 8, 1974 voted to require that, to be considered at the annual business meeting, proposed resolutions must be submitted to the Secretary at least one month in advance in writing with signatures of the proposer and the second, both of whom must be members in good standing. The Secretary will reproduce the proposed resolutions and make copies available in advance of the meeting. The next business meeting will begin at 9:30 p.m. on September 6, 1980. The deadline for proposed resolutions is accordingly August 6. They should be sent to the Secretary, American Economic Association, 1313 21st Avenue, So., Suite 809, Nashville, TN 37212.

Nominations for AEA Officers: 1981

The Electoral College on March 21 chose Gardner Ackley as nominee for President-Elect of the American Economic Association in the balloting to be held in the autumn of 1980. Other nominees (chosen by the 1980 Nominating Committee) are: Vice-President (two to be elected), Otto Eckstein, Leonid Hurwicz, Dale W. Jorgenson, and Alice Rivlin; for members of the Executive Committee (two to be elected), Elizabeth Bailey, Robert J. Gordon, Jacob Mincer, and Sherwin Rosen.

Under a change in the bylaws as described in the *Papers and Proceedings* of this *Review*, May 1971, page 472, additional candidates may be nominated by petition, delivered to the Secretary by August 1, including signatures and addresses of not less than 6 percent of the membership of the Association for the office of President-Elect, and not less than 4 percent for each of the other offices. For the purpose of circulating petitions, address labels will be made available by the Secretary at cost.

1981 Nominating Committee of AEA

In Accordance with Section IV, paragraph 2, of the bylaws of the American Economic Association as amended in 1972, President-Elect William Baumol has appointed a Nominating Committee for 1981 consisting of Carl F. Christ, Chair; Marcus Alexis, Robert Dorfman, Juanita Kreps, Lester Lave, Lester Thurow, and Robert Willig. Attention of members is called to the part of the bylaw reading, "In addition to appointees chosen by the President-Elect, the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than 2 percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the membership immediately following its appointment and the membership invited to suggest nominees for the various officers to the Committee."

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada, that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to air fare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for applications to be received in the ACLS office are: meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 800 Third Avenue, New York, NY 10022, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting. Even when plans are incomplete, a prospective applicant should request forms in advance of the cut-off date, since deadlines are firm and no exceptions are permitted. Awards will be announced approximately two months after each deadline.

New Journal–Call for Papers: The Western Tax Association, together with Washington State University, announces its new refereed journal, *Western Tax Review*. The *WTR* will deal with all aspects of public finance, but especially with major questions of public policy. Initially, it will appear once a year. To submit a paper, send three copies to Professor Reed R. Hansen, Editor, College of Economics and Business, Washington State University, Pullman, WA 99163. The submission fee is $30 for nonmembers and $20 for members, half of which will be refunded for papers not accepted.

The International Institute of Public Finance will hold its thirty-sixth World Congress of Public Finance in Jerusalem, Israel, August 25-29, 1980. The topic will be "Public Finance and Public Employment" (in market and socialist economies). For more information, contact Professor Horst Claus Recktenwald, President of IIPF, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lange Gasse 20, 8500 Nürnberg, Germany. The World Congress of Public Finance in September 1981 will be held in Tokyo, Japan.

Twelve long-term (six to ten months) and nine short-term (two to three months) research awards, without restriction as to field, are offered for 1981–82 by the Indo-U.S. Subcommission on Education and Culture. Applicants must be U.S. citizens at the postdoctoral or equivalent professional level. The fellowship program seeks to open new channels of communication between academic and professional groups in the United States and India, and to encourage a wider range of research activity between the two countries than has previously existed. Therefore, scholars and professionals who have limited or no experience in India are especially encouraged to apply. *The application deadline is July 1, 1980.* Forms and further information are available from the Council for International Exchange of Scholars, Attention: Indo-American Fellowship Program, Eleven Dupont Circle, Washington, DC 20036, telephone 202+833-4978.

The International Union for the Scientific Study of Population (IUSSP) is a nonprofit, scientific organization founded in 1928. Its primary objective is the advancement of the sciences of demography. Conferences on a full range of topics are organized by the Union every four years; smaller conferences of interest to specialists or of particular relevance to particular regions are held between the quadrennial general meetings. The 1977 general meeting was in Mexico; the 1981 meeting will be held in Manila. There are currently seven scientific committees and four working groups of the Union that stimulate research and exchange ideas in areas that seem inadequately developed, primarily by means of holding seminars on specific topics. The Union publishes Proceedings of its meetings and of the seminars sponsored by its committees, and a series of IUSSP papers. Conference Proceedings are distributed to all registered participants and are available to members at a discount price. Members also receive the IUSSP papers and are provided with subscriptions to four specialized journals: *Population* (Paris); *Population Studies* (London); *Genus* (Rome); *Population Index* (Priceton). Individuals interested in joining the Union may request nomination forms from the Executive Secretary, IUSSP, 5 rue Forgeur, 4000 Liège, Belgium.

The third International Time-Series Meeting (first American conference) will be held August 14-15, 1980 at the Shamrock Hilton Hotel, Houston, Texas. The conference will convene immediately following the American Statistical Association meetings at the same location. For further information, contact M. Ray Perryman, Director, Center for the Advancement of Economic Analysis, Hankamer School of Business, Baylor University, Waco, TX 76706.

Call for Papers: The Rocky Mountain Conference on British Studies will hold its annual meeting October 10-11, 1980 at Fort Lewis College, Durango, Colorado. Submit paper proposals by July 1 to Stanley Palmer, Department of History, University of Texas, Arlington, TX 76019. For general information, contact Larry Gasser, Department of English, Fort Lewis College, Durango, CO 81301.

Call for Papers: The Conference on Research in Income and Wealth will hold a conference on the Measurement of Wages in the fall, 1981. The conference's focus will be on total labor compensation (including traditional fringes, hiring, and training costs, and costs associated with nonpecuniary benefits) as an input cost to employers. Papers are invited which present empirical research on any of these subjects. Authors should also be prepared to draw implications from their research results for the kinds of data that should be collected or for the way labor cost data should be measured. Proposals should be submitted to Jack E. Triplett, Assistant Commissioner for Research and Evaluation, Rm 2021, U.S. Bureau of Labor Statistics, 441 G St., NW, Washington, DC 20212.

The Midwest Economics Association will hold its next meeting April 3-5, 1981 at the Galt House in Louisville, Kentucky. Program suggestions should be sent *now* to Professor Burton Weisbrod, MEA President, Department of Economics, University of Wisconsin, Madison, WI 53706. For general MEA information, contact Professor Allen F. Larsen, MEA Secretary-Treasurer, Department of Economics, St. Cloud State University, St. Cloud, MN 56301.

The World Rehabilitation Fund International Exchange of Information in Rehabilitation announces the availability of a monograph: *The Value of Independent Living: Looking at Cost Effectiveness in the U.K.* (Issues for Discussion in the U.S.), written by Jean Simkins, The Economist Intelligence Unit, London. This monograph has been written and published under a rehabilitation services administration grant which addresses itself to the filling of information gaps in the United States by arranging for foreign experts to write monographs in areas of concern. To obtain a copy, write Ms. Diane Woods, Dissemination, World Rehabilitation Fund, 400 East 34th St., New York, NY 11016.

Members of the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics are pleased to announce the institution of an annual Leonard J. Savage Award of five hundred dollars ($500) for an outstanding doctoral dissertation in the area of Bayesian Econometrics and Statistics. To be considered for the 1980 Savage Award, a doctoral dissertation must be submitted by the dissertation supervisor before July 1, 1980 and accompanied by a short letter from the supervisor summarizing the main results of the dissertation. Dissertations completed after January 1, 1977 are eligible to be considered for the 1980 Savage Award. An Evaluation Committee will be appointed by the board of the Leonard J. Savage Memorial Trust Fund (M. H. DeGroot, S. E. Fienberg, S. Geisser, J. B. Kadane, E. E. Leamer, J. W. Pratt and A. Zellner, chairman) to evaluate dissertations that are submitted for the Savage Award. Dissertations and supporting letters should be sent to Professor Arnold Zellner, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637.

Sloan Research Fellowships in Economics: Nominations are now being accepted for six fellowships to be awarded in 1981 to young faculty members (no more than 32 years of age). Direct applications are not accepted. The deadline for nominations is September 15, 1980. For full information and nomination forms, contact Sloan Research Fellowships, Alfred P. Sloan Foundation, 650 Fifth Avenue, New York, NY 10020.

The Sixth Symposium on Statistics and the Environment will be held at the National Academy of Sciences on October 6-7, 1980. The Symposium provides a forum for policymakers, scientists, and people affected by environmental decisions. It will focus on how statistics, economics, epidemiology, toxicology, and other scientific bodies of information are used in environmental decisions by administrative agencies, legislatures, and the courts. Those interested should contact the AEA representative to the Steering Committee for the Sixth Symposium, Eugene Seskin, Bureau of Economic Analysis, BE-62A, 1401 K Street, NW, Washington, D.C. 20230 (telephone 202+523-0734).

Symposium on Economic and Social Value of Weather and Climate Information, January 28-29, 1981, San Diego: Call for Paper: Cosponsored by the American Economic Association, the American Agricultural Economic Association, and the American Meteorological Society as part of the AMS annual meeting, the objectives are to describe methods of measuring meteorological information value and to present results of case studies. Those working in this area are invited to submit a detailed one-page abstract no later than August 1, 1980 to Dr. Allen H. Murphy, Department of Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331. Selections will be made by the Program Committee consisting of Murphy for AMS, Charles W. Howe for the AEA, and Ludwig Eisgruber for AAEA, and will be announced September 15th.

Death

Bobby L. Hamm, professor of economics, Louisiana Tech University, Nov. 18, 1979.

Retirements

George H. Cochran, associate professor, department of economics, Kent State University, June 15, 1979.

Finn B. Jensen, professor of economics, Lehigh University, June 1979.

Vito Natrella, director, Statistics Division, Internal Revenue Service, Jan. 5, 1980.

Ernest F. Patterson, professor of economics, Davidson College, June 1979.

L. Reed Tripp, professor of economics, Lehigh University, June 1979.

William J. Weiskopf, professor, department of economics, Kent State University, Jan. 1, 1980.

Visiting Foreign Scholars

Richard T. Baillie, University of Aston, England: visiting associate professor, department of economics, Wayne State University, 1979-80.

Alfred R. Endres, Konstanz, West Germany: visiting assistant professor, department of economics, State University of New York-Buffalo, Sept. 1, 1979–July 1, 1980.

James Harris, University of Singapore: visiting assistant professor, department of economics, Wayne State University, 1979-80.

Michael Landsberger, Haifa, Israel: visiting professor, department of economics, State University of New York-Buffalo, summer 1980.

Patrick McMahon, University of Birmingham, England: visiting associate professor, department of economics, Wayne State University, spring 1980.

Promotions

Raj Aggarwal: professor of finance and international business, University of Toledo, 1979.

Jeffrey I. Chapman: associate professor of public administration, School of Public Administration, University of Southern California, Sept. 1, 1979.

James A. Hanson: professor of economics, Brown University, July 1, 1980.

John R. McNamara: professor of economics, Lehigh University, Sept. 1, 1979.

Arturo C. Porzecanski: associate economist, Morgan Guaranty Trust Company of New York, Nov. 1, 1979.

Administrative Appointments

Edward B. Bell: chairman, department of economics, Cleveland State University, July 1979.

Paul H. Engelmann: chairman, department of economics, Central Missouri State University, Sept. 1, 1979.

Alvin N. Puryear: vice president for organization and management, Ford Foundation, Feb. 1, 1980.

Leon M. Schur: acting chancellor, University of Wisconsin-Milwaukee, Sept. 1, 1979.

Arthur W. Wright: professor and head, department of economics, University of Connecticut, Sept. 1, 1979.

Appointments

Frances E. Altvatter, University of Pennsylvania: assistant professor of economics, Bryn Mawr College, Sept. 1979.

William R. Bowman, Goucher College: assistant professor, economics department, U.S. Naval Academy, Aug. 1979.

Kristine L. Chase, University of Maryland: visiting assistant professor, economics department, U.S. Naval Academy, Aug. 1979.

Deborah J. Danker: economist, international research department, Industrial Economies Division, Federal Reserve Bank of New York, Dec. 3, 1979.

W. Davis Dechert: assistant professor of economics, State University of New York-Buffalo, Sept. 1, 1979.

Frank T. de Vyver: visiting professor, department of economics, Davidson College, fall term 1979-80.

James A. Duffield: research associate, department of economics, Iowa State University, Oct. 19, 1979.

Thomas J. Espenshade, Florida State University: senior research associate, Women and Family Policy Research Program, The Urban Institute, Jan. 1980.

Gilbert R. Ghez, University of Illinois-Chicago Circle: associate professor of management and economics, Roosevelt University, Sept. 1, 1979.

William S. Haraf: assistant professor of economics, Brown University, July 1, 1979.

Samuel P. Hayes: staff consultant, Committee on Agriculture, U.S. House of Representatives, July 19, 1979.

Thomas J. Hyclak: assistant professor of economics, Lehigh University, Sept. 1, 1979.

Laurence G. Kantor: instructor, economics department, Lehigh University, Jan. 1, 1980.

Stephen Karlson, University of Wisconsin: assistant professor, department of economics, Wayne State University, fall 1979.

Daniel Kohler, University of Michigan: assistant professor, department of economics, Wayne State University, fall 1979.

Ralph Kolinski, University of Windsor, Canada: visiting lecturer, department of economics, Wayne State University, 1979-80.

Richard D. Marcus, University of Chicago: instructor, department of economics, Cleveland State University, Sept. 1979.

Michael J. Monson: research associate, department of economics, Iowa State University, Oct. 15, 1979.

Kazuo Nishimura: assistant professor of economics, State University of New York-Buffalo, Sept. 1, 1979.

Clark G. Ross: assistant professor, department of economics, Davidson College, Sept. 1979.

David J. Salant: assistant professor of economics, State University of New York-Buffalo, Sept. 1, 1979.

James S. Shortle: research associate, department of economics, Iowa State University, Nov. 1, 1979.

W. Tate Whitman: visiting professor, department of economics, Davidson College, winter term 1979-80.

Lee D. Zinser: research associate, department of economics, Iowa State University, Oct. 15, 1979.

Leaves for Special Appointment

Robert L. Avinger, Davidson College: visiting research fellow, Institute of United States Studies; and academic visitor, London School of Economics, 1979-80.

Alfred J. Francfort, University of Pittsburgh-Johnstown: economist, Energy and Minerals Division, U.S. General Accounting Office, Aug. 1979–Aug. 1980.

Roger D. Little, U.S. Naval Academy: visiting professor, U.S. Naval Postgraduate School, Monterey, California, Sept. 1979.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1980 issue of the *Review* will carry the seventy-seventh list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1980. This announcement is an invitation to send us information for the preparation of the list. This announcement supersedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1980 issue the names of those who will have been awarded the doctoral degree since June 1979 and the titles of their dissertations. Dissertation abstracts will no longer be published, as these are published elsewhere.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ________

Name: LAST NAME IN CAPS: First Name, Initial ________

Institution Granting Degree: ________

Degree Conferred (Ph.D. or D.B.A.) ________ Year ________

Dissertation Title: ________

BALLINGER PUBLISHING COMPANY

CAPITAL INVESTMENT AND SAVING
Editor, George M. von Furstenberg, *Indiana University*
(Sponsored by the American Council of Life Insurance)
Vol. I: Social Security Versus Private Saving
January 1980 $35.00 420 Pages
Vol. II: The Government and Capital Formation
March 1980 $35.00 554 Pages
Vol. III: Capital, Efficiency, and Growth
March 1980 $35.00 584 Pages

THE FUNCTIONING OF FLOATING EXCHANGE RATES
Theory, Evidence, and Policy Implications
Editors, David Bigman and Teizo Taya, *International Monetary Fund*
May 1980 ca.$25.00 ca.432 Pages

BUSINESS CYCLES, INFLATION, AND FORECASTING
Geoffrey H. Moore, *National Bureau of Economic Research*
May 1980 ca.$35.00 ca.432 Pages

SCHOOL FINANCE POLICIES AND PRACTICES
The 1980s: Decade of Conflict
(The 1980 Yearbook of the American Education Finance Association)
Editor, James W. Guthrie
Summer 1980 Pages and price to be announced.

LAGGING PRODUCTIVITY GROWTH
Causes and Remedies
Schlomo Maital and Noah Meltz
May 1980 Pages and price to be announced.

DEMAND REVELATION AND THE PROVISION OF PUBLIC GOODS
Edward H. Clarke, *Smithsonian Institution*
Spring 1980 ca.$22.50 ca.356 Pages

Telephone: (617) 492-0793

COMMODITY MARKETS AND LATIN AMERICAN DEVELOPMENT
A Modeling Approach
Editors, Walter C. Labys, M. Ishaq Nadiri, and Jose Nunez del Arco
April 1980 ca.$25.00 ca.278 Pages

PUBLIC EMPLOYMENT AND STATE AND LOCAL GOVERNMENT FINANCE
Editors, Roy W. Bahl, Jesse Burkhead, and Bernard Jump, Jr., *Syracuse University*
April 1980 ca.$25.00 ca.264 Pages

TELECOMMUNICATIONS DEMAND
A Survey and Critique
Lester D. Taylor, *University of Arizona*
April 1980 ca.$27.50 ca.240 Pages

RISK-BENEFIT ANALYSIS
Richard Wilson and Edmund Crouch, *Harvard University*
Spring 1980 ca.$22.50 ca.304 Pages

SUPPORTED WORK
Summary and Findings of the National Supported Work Demonstration
The Board of Directors, *Manpower Demonstration Research Corporation*
April 1980 Price to be announced. 192 Pages

TAX AND INVESTMENT POLICIES FOR HARD MINERALS
Public and Multinational Enterprises in Indonesia
S. Malcolm Gillis and Ralph E. Beals, *Harvard Institute for International Development*
April 1980 ca.$30.00 ca.320 Pages

STALEMATE IN TECHNOLOGY
Gerhard Mensch, *International Institute of Management, Berlin*
1979 $18.50 280 Pages

ECONOMIC DISINCENTIVES FOR ENERGY CONSERVATION
Joe W. Russell, *Environmental Law Institute*
1979 $16.00 160 Pages

BALLINGER PUBLISHING COMPANY
17 Dunster Street, Cambridge, MA 02138

New books from the World Bank

World Tables 1980

from the data files of the World Bank

One of the most complete collections of time series data ever assembled, this volume contains nearly 600 pages of tables analyzing various economic and social indicators for most of the countries of the world. $27.50 hardcover, $10.95 paperback

The Economics of Power System Reliability and Planning

Theory and Case Study

Mohan Munasinghe

This is the first completely integrated treatment of electric power system reliability. The author focuses on the economic losses incurred by consumers because of power failures and develops a model that more accurately determines the true costs of electric power. $24.00 hardcover; $8.95 paperback

Farm Budgets

From Farm Income Analysis to Agricultural Project Analysis

Maxwell L. Brown

This book clarifies the relation between simple farm income analysis and the broader field of agricultural project analysis and illustrates how the transition is made from one to the other. *A World Bank Staff Occasional Paper* $15.00 hardcover; $6.50 paperback

Malaysia

Growth and Equity in a Multiracial Society

Kevin Young, Willem Bussink, and Parvez Hasan

This report examines the performance of the Malaysian economy since 1960, discusses current issues, and assesses the outlook for the future with special attention given to the government's efforts to reduce poverty and redress racial imbalances in the economy. *A World Bank Country Economic Report* $25.00 hardcover; $7.95 paperback

Confronting Urban Malnutrition

The Design of Nutrition Programs

James E. Austin

Presupposing a political commitment to meet basic human needs in the urban setting, this study analyzes the systematic planning, design of interventions, program management, and community action that is necessary to translate political will into improved nutrition for the urban poor. *A World Bank Staff Occasional Paper* $6.50 paperback

A Model of an Agricultural Household

Theory and Evidence

Howard N. Barnum and Lyn Squire

In this innovative study, an economic analysis of agricultural households in Malaysia, the authors show that the integration of production and consumption decisions yields substantially different policy conclusions from those obtained by the traditional approach: analysis of production or consumption alone. *A World Bank Staff Occasional Paper* $6.95 paperback

published by

Johns Hopkins

The Johns Hopkins University Press, Baltimore, Maryland 21218

NORTH-HOLLAND ANNOUNCES NEW BOOKS·NEW BOOKS·NEW BOOKS·NEW BOOKS·NEW

NORTH-HOLLAND PUBLISHING COMPANY
52 Vanderbilt Ave., New York, N.Y. 10017/ P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Bayesian Analysis in Economic Theory and Time Series Analysis

The 1977 Savage Dissertation Award Theses

Studies in Bayesian Econometrics, Vol. 2

By CHARLES A. HOLT, Jr., Univ. of Minnesota, and ROBERT W. SHORE, Westinghouse R&D, Pittsburgh

1980. 192 pages.
US $39.00/Dfl. 80.00
ISBN 0-444-85414-2

The 1977 Leonard J. Savage prize was awarded to Charles A. Holt, Jr., for his thesis, "Bidding for Contracts", while Robert W. Shore's thesis, "A Bayesian Approach to the Spectral Analysis of Stationary Time Series", was accorded honorable mention. In view of the high quality of both theses, it was decided to publish these together in this volume.

Socioeconomic Success

A Study of the Effects of Genetic Endowments, Family Environment, and Schooling.

Contributions to Economic Analysis, Vol. 128

By JERE R. BEHRMAN, Univ. of Pennsylvania, ZDENEK HRUBEC, National Research Council, PAUL TAUBMAN, Univ. of Pennsylvania, and TERENCE J. WALES, Univ. of British Columbia

1980. 290 pages.
US $58.50/Dfl. 120.00
ISBN 0-444-85410-X

A large sample of white male twins is used to estimate a model of schooling and occupational attainment and earnings at about age 50. Methods are developed and used to control for unobserved differences in genetic endowments and family environment when estimating the relationship of schooling to earnings. Methods are also developed and used to partition the variance of earnings, etc. into genetics, common environment and noncommon environment components.

Theories of Urban Externalities

Studies in Regional Science and Urban Economics, Vol. 6

By YOSHITSUGU KANEMOTO, Univ. of British Columbia, Vancouver

1980. 298 pages.
US $31.75/Dfl. 65.00
ISBN 0-444-85483-5

In this volume several aspects of externalities in cities are analyzed using extensions of a standard residential land use model. Topics covered are: optimal and market city sizes, local public goods, traffic congestion, externalities between different types of households, and the growth of a system of cities.

Statistical Methods of Econometrics

Third Revised Edition

Studies in Mathematical and Managerial Economics, Vol. 6

By E. MALINVAUD, Institut National de la Statistique et des Etudes Economiques, Paris

1980. 760 pages.
US $48.75/Dfl. 100.00
ISBN 0-444-85473-8

"... my respect for this book is virtually unbounded. Given a class of mathematically knowledgeable graduate students in econometrics, the book has no peer."
Economica

"... this textbook stands in a class of its own and is bound to have a strong influence on the teaching of econometrics for many years to come."
The Economic Journal

"... this book is at once more comprehensive in coverage and more advanced in treatment than any other econometric text... His success in this task has earned M. Malinvaud the gratitude of his readers."
Journal of the Royal Statistical Society

0727 NH

New from Columbia

Essays on the Theory of Joint Production

LUIGI PASINETTI, EDITOR. This volume includes a formalization of Piero Sraffa's general scheme of joint production; the formulation of new tools for economic analysis in the vertically integrated sectors; an analysis of the many problems connected with durable means of production; an inquiry into the effects of some forms of technical progress; and a formalization of Sraffa's scheme for non-produced means of production.

256 pages, $30.00

Dependency Approaches to International Political Economy

A Cross-National Study

VINCENT A. MAHLER. Dependency theory has recently attracted considerable attention as an important, if highly controversial, perspective on problems of underdevelopment in the Third World. Vincent A. Mahler's study examines the dependency theorists' most important hypotheses in light of data derived from a cross-national study of 70 less-developed countries.

232 pages, $17.50

The Measurement of Economic Growth

DAN USHER. This unique study explores in a single volume topics ordinarily treated separately in economic literature: the conceptual problems of defining and measuring growth and the practical problems of organizing data. Emphasizing both theory and application to real-world cases, the author raises many key issues in the controversy over the importance of economic growth.

336 pages, $22.50

The Economics of Interdependence

Economic Policy in the Atlantic Community;
Published for the Council on Foreign Relations

RICHARD N. COOPER. Now back in print, this volume offers a penetrating analysis of the complex and often contradictory economic currents flowing between the two hemispheres. Richard N. Cooper addresses the central problem facing the world's macroeconomists: how to preserve the manifold benefits accruing from international trade while maintaining sufficient domestic freedom to pursue attractive political and economic objectives. *A Morningside Book*

316 pages, $20.00 cloth, $6.50 paper (tent.)

To order send check or money order to Dept. JN at the address below, including $1.30 per order for postage and handling.

COLUMBIA UNIVERSITY PRESS
Address for orders:
136 South Broadway, Irvington, New York 10533

A Guide to Post-Keynesian Economics

EDITED BY ALFRED S. EICHNER
WITH A FOREWORD BY JOAN ROBINSON

"An informed, technically competent, yet wholly readable set of essays on economics after Keynes. A book that will help modernize both teaching and teachers."
—**John Kenneth Galbraith, Harvard University**

"For anyone who wishes to know what is going on in the world of economic thought, it is, I believe, indispensable."
—**Robert L. Heilbroner in *The New York Review of Books***

Cloth $12.95 Paper $4.95

This first comprehensive statement of the post-Keynesian position includes:

MACRODYNAMICS *John Cornwall*
PRICING *Peter Kenyon*
INCOME DISTRIBUTION *J. A. Kregel*
TAX INCIDENCE *A. Asimakopulos*
PRODUCTION THEORY *Richard X. Chase*
THE SRAFFIAN CONTRIBUTION *Alessandro Roncaglia*
THE LABOR MARKET *Eileen Appelbaum*
MONETARY FACTORS *Basil J. Moore*
THE INTERNATIONAL DIMENSION *John B. Burbidge*
NATURAL RESOURCES *Paul Davidson*
A LOOK AHEAD *Alfred S. Eichner*

M. E. Sharpe, Inc. 901 N. BROADWAY, WHITE PLAINS, N.Y. 10603

M. E. Sharpe, Inc., 901 North Broadway, White Plains, NY 10603

Please send_____copies of A GUIDE TO POST-KEYNESIAN ECONOMICS in cloth @$12.95.
Please send_____copies in paperback @$4.95.
My check for $_______is enclosed. (Publisher pays postage and handling if payment accompanies order.)

Please charge my ☐Master Charge ☐VISA (check one)
Account Number____________________ Exp. date__________
Signature ____________________________
☐Please send your current brochure in Economics.

NAME ____________________________
AFFILIATION ________________________
ADDRESS __________________________
CITY__________________STATE & ZIP__________

National Bureau
of Economic
Research

Three New Titles from the

Rational Expectations and Economic Policy

Edited by Stanley Fischer

Cloth 312 pages $22.00

The Measurement of Capital

Edited by Dan Usher

Studies in Income and Wealth, vol. 45

Cloth 560 pages $40.00

New Developments in Productivity Measurement and Analysis

Edited by John W. Kendrick
and Beatrice N. Vaccara

Cloth 768 pages $52.00

And

The System-Wide Approach to Microeconomics

Henri Theil

Cloth 288 pages $23.00

The University of Chicago Press Chicago 60637

PRAEGER
**entering our fourth
decade of
distinguished publishing**

MANAGING GROWTH IN THE 1980's
Toward a New Economics
By **Robert D. Hamrin**, Environmental Protection Agency
Challenging conventional economic wisdom, Hamrin suggests that the traditional sources of economic growth have become non-supportive of growth, and examines the forces that are most critical in shaping the economy. He concludes that the new realities governing economic performance demand a wholesale rethinking of economics. Finally, he suggests new actions and policies, shaped by a selective growth ethic, aimed at human resource development, and formed within the framework of a future-oriented national growth policy.
318 pp. 1980 $23.95 LC 79-19459 ISBN 0-03-054061-5
Text edition: $9.95 ISBN 0-03-054056-9

THE MONETARY APPROACH TO INTERNATIONAL ADJUSTMENT
Edited by **Bluford H. Putnam**, Chase Manhattan Bank, and **D. Sykes Wilford**, Chase Manhattan Bank
Here's an historical/developmental approach ideal for students with diverse academic backgrounds. With contributors drawn from commercial banks, the Fed, and academia, this essay approach combines with a 25-page bibliography to lay the groundwork for advanced research by upper-level students. The co-authors refine the variables in the framework model and in the statistical tools used to test the model's implication. The result is a clear picture of how the monetary approach evolved, its current state, and the most promising avenues for future research.
299 pp. 1979 $18.95 LC 78-19753 ISBN 0-03-046711-X
Text edition: $8.95 ISBN 0-03-056211-2

Praeger Publishers
521 Fifth Avenue, New York, NY 10017

Leadership...

HARPER & ROW LEADS THE ECONOMICS FIELD WITH THE BEST INTRODUCTORY BOOKS

☑ **Waud—ECONOMICS**
Combined Volume: ISBN 0-06-046963-3. 848 pages.
MACROECONOMICS, ISBN 0-06-046965-X. 478 pages.
MICROECONOMICS, ISBN 0-06-046966-8. 423 pages.

☑ **Miller—ECONOMICS TODAY,** ***Third Edition***
Combined Volumes: ISBN 0-06-385481-3. 747 pages.
ECONOMICS TODAY: THE MACRO VIEW, Third Edition
ISBN 0-06-385488-0. 432 pages.
ECONOMICS TODAY: THE MICRO VIEW, Third Edition
ISBN 0-06-385489-6. 433 pages.

☑ **Lipsey & Steiner—ECONOMICS,** ***Fifth Edition***
Combined Volume: ISBN 0-06-044016-3. 982 pages.
MACROECONOMICS, Fifth Edition, ISBN 0-06-043976-9. 513 pages.
MICROECONOMICS, Fifth Edition, ISBN 0-06-043981-5. 512 pages.

☑ **Hunt & Sherman—ECONOMICS**
An Introduction to Traditional and Radical Views, ***Third Edition***
ISBN 0-06-043017-6. 736 pages. Paper.

☑ **North & Miller—THE ECONOMICS OF PUBLIC ISSUES,** ***Fifth Edition***
ISBN 0-06-044875-X. 214 pages. Paper.

☑ **Elzinga—ECONOMICS: A READER**
Third Edition
ISBN 0-06-04192-1. 304 pages. Paper.

To request examination copies, write to Harper & Row, 10 East 53d Street, New York, N.Y. 10022. Suite 5D (888). Please include course title, enrollment and present text.

Harper & Row
New York • San Francisco

AEA sponsored Group Life Insurance for you and your family—at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA has joined a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by dividend credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 45% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future dividends and credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program B-3
1707 L Street, N.W.—Suite 700
Washington, D.C. 20036

Please send me more information about the AEA Life Insurance Plan.

Name ____________________ Age _____

Address ____________________

City ____________ State ______ Zip ______

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Whodunit?
Wedunit!!

THOMAS HORTON AND DAUGHTERS has published the first mystery novel to be used widely in undergraduate teaching. Over 200 adoptions and the list is growing.

Economists have enjoyed matching wits with Professor Henry Spearman. Now your students have the opportunity to learn—and the evidence indicates enjoy learning – economic principles from MURDER AT THE MARGIN by Marshall Jevons.

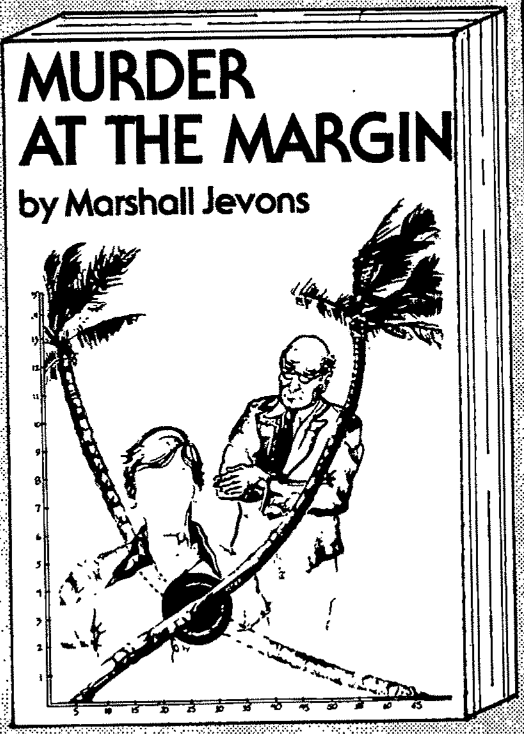
Priced at $3.95 in paperback and $7.95 hardcover. TEST QUESTIONS AVAILABLE FROM PUBLISHER.

As the book review in the WALL STREET JOURNAL declared:

"IF THERE IS A MORE PAINLESS WAY TO LEARN ECONOMIC PRINCIPLES, SCIENTISTS MUST HAVE RECENTLY DISCOVERED HOW TO IMPLANT THEM IN ICE CREAM."

"I thought the economic argument extremely ingenious and the idea of using economic analysis as a way to solve the mystery most original."
Milton Friedman

"At last, a new kind of mastermind—a rational 'homoeconomicus' and libertarian. If Henry Spearman had not existed, God would have had to invent him. Marshall Jevons did, to his readers' benefit.
Paul A. Samuelson



THOMAS HORTON & DAUGHTERS
22 Appleton Place, Box 3
Glen Ridge, New Jersey 07028

Begin making plans to attend the

Ninety-Third Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

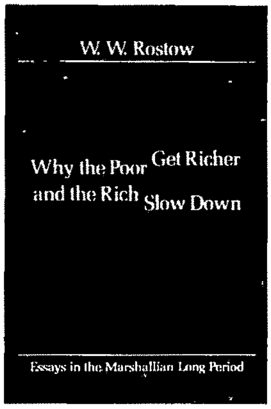
DENVER, COLORADO

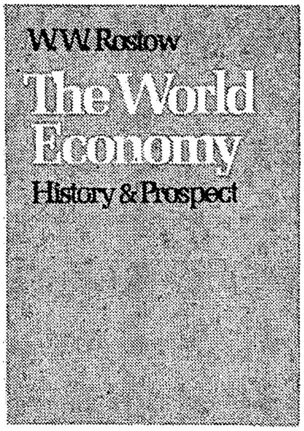
September 5-7, 1980

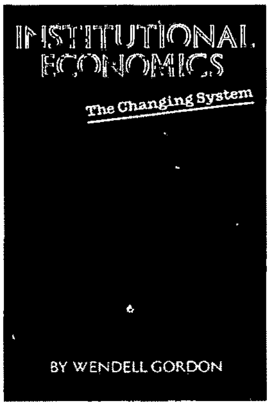
The 1980 Professional Placement Service will be held December 28-30 at the Dallas Hilton, Dallas, Texas. Please note that there will not be a Professional Placement Service in Denver at the September meetings.

See the Notes section of the June *AER* for the American Economic Association's preliminary program.

The 1981 meeting will be held in Washington, DC, December 28-30.


W. W. Rostow
Why the Poor Get Richer and the Rich Slow Down
Essays in the Marshallian Long Period





Why the Poor Get Richer and the Rich Slow Down
ESSAYS IN THE MARSHALLIAN LONG PERIOD
By W. W. Rostow
Bearing on fundamental issues of economic theory, history, and public policy, this volume elaborates and goes beyond themes enunciated in W. W. Rostow's previous works. The eight essays presented here are unified by the author's insistence that neo-Keynesian and neoclassical theory are an inadequate basis for economic analysis and policy prescription. 456 pages; $19.95

The World Economy
HISTORY AND PROSPECT
By W. W. Rostow
"The scope of the book is astonishing. . . . it traces the course of the world economy since the 18th century."
—Lindley H. Clark, Jr., *Wall Street Journal*
". . . the path of evolution from the pre-industrial to the advanced industrial condition is meticulously chronicled."
—Jock Bruce-Gardyne, *Daily Telegraph*
1978, 877 pages, 275 tables, 157 charts; $34.50

Institutional Economics
THE CHANGING SYSTEM
By Wendell Gordon
This volume clearly and evocatively sets out the philosophy of economic institutionalism, which, in Wendell Gordon's view, offers the most hope for understanding and solving the economic problems of the twentieth century. Gordon's argument that the dynamics of technical change, the institutionalism of behavior norms, human biology, and the resource endowment all interact in an ongoing social process is a central tenet of the institutional approach.
400 pages, 15 figures, 4 tables; $9.95, paper; $19.95, cloth

Toward a Reasonable Society
By C. E. Ayres
"Professor Ayres provides in this book a rigorous account of our world. . . . By appealing wholeheartedly to the scientific-technological bases of life, [he] develops a sanguine philosophy, liberal and liberating." —*Technology and Culture*
1961, 309 pages; $7.95, paper

Science and Ceremony
THE INSTITUTIONAL ECONOMICS OF C. E. AYRES
Edited by William Breit and William Patton Culbertson, Jr.
". . . *Science & Ceremony* has reawakened interest in one of the most striking characters in the economics profession of the United States in this century." — Kenneth E. Boulding
1976, 228 pp.; $14.95

071 TEXAS

To order directly send payment to: **University of Texas Press**
Texas residents add 5% sales tax. POST OFFICE BOX 7819 AUSTIN, TEXAS 78712

ANNOUNCING

1980 New Edition

Guide to Graduate Study in Economics and Agricultural Economics

in the United States of America and Canada

Designed to provide students anticipating graduate study in economics and agricultural economics, and their advisors, with information on available graduate training programs.

Includes descriptions of 262 graduate programs, supplemented by comparative data and information for prospective students, domestic and foreign.

PUBLISHED BY THE ECONOMICS INSTITUTE,
University of Colorado at Boulder, Boulder, Colorado 80309 under the auspices of the **American Economic Association** and the **American Agricultural Economics Association.**
PRICE: $15.50 per copy

ORDER FORM

RICHARD D. IRWIN, INC.
1818 Ridge Road
Homewood, Illinois 60430

Please send me ________ copies of *Guide to Graduate Study in Economics and Agricultural Economics in the United States of America and Canada,* 5th edition.

Enclosed is my check/money order for $________ ($________ per copy).

Please print or type:

Name ____

Address ____

City ______________ State ____________ Zip ______________

STAFF PAPERS

Published by the International Monetary Fund in March, June, September, and December. Summaries of each paper in French and Spanish, as well as in English.

Vol. 27 March 1980 No. 1

Distributional Aspects of Stabilization Programs in Developing Countries
OMOTUNDE JOHNSON and JOANNE SALOP

A Supply Framework for Exchange Reform in Developing Countries: The Experience of Sudan .... KARIM NASHASHIBI

The Optimal Basket in a World of Generalized Floating
LESLIE LIPSCHITZ and V. SUNDARARAJAN

Why Does the Current Account Matter?
JOANNE SALOP and ERICH SPITÄLLER

Estimation of the Timing Asymmetry in International Trade
WILLIAM L. HEMPHILL

The Influence of Social Security on Household Savings: A Cross-Country Investigation
GEORGE KOPITS and PADMA GOTUR

New prices, effective June 1, 1980, are: Subscription: US$9.00 a volume; US$3.00 a single issue. Special rate to university libraries, faculty, and students: US$4.00 a volume; US$1.00 a single issue.

Advice on payment in other currencies will be given upon receipt of order.

The Secretary, Attention: Publications Section, International Monetary Fund, Washington, D.C. 20431 U.S.A.

ALLIED SOCIAL SCIENCE ASSOCIATIONS

1980 ANNUAL MEETINGS • DENVER • SEPTEMBER 5-7

Preregistration material for the 1980 annual meetings of the Allied Social Science Associations, to be held in Denver, has been mailed. If you have not received a packet of information and plan to attend the meetings, please complete the form below and return it to:

American Economic Association
Attn: Melissa Williams
1313 21st Avenue South
Nashville, TN 37212

Name in full____________________________________
First Middle Last

Address____________________________________

____
City State/Country Zip/Post Code

An indispensable tool for:

Economists Teachers Students

Researchers Authors

Reference Librarians

Index of Economic Articles

prepared under the auspices of

The Journal of Economic Literature

of the

American Economic Association

Each volume in the *Index* lists, both by subject and by author, articles in major economic journals and in collective volumes published during a specific year. The *Index* volumes also include articles or testimony from selected congressional hearings or government documents published during the year. Many of the features in the *Index* cannot be found in any other single reference source.

Currently available are:

Volume	Year Covered
XI	1969
XII	1970
XIII	1971
XIV	1972
XV	1973
XVI	1974
XVII	1975

Future volumes will be published regularly to keep the series as current as possible.

***Price:* $35.00 per volume (special 30% discount to A.E.A. members)**

Distributed by

RICHARD D. IRWIN, INC.

Homewood, Illinois 60430

Note: Volumes I through X, VI-A, and VII-A, previously published by Irwin, are available at $7.95 per copy.

The American Economic Review

ARTICLES

W. A. LEWIS — The Slowing Down of the Engine of Growth

J. L. STEIN — The Dynamics of Spot and Forward Prices in an Efficient Foreign Exchange Market

V. L. SMITH — Experiments with a Decentralized Mechanism for Public Good Decisions

M. AOKI — A Model of the Firm as a Stockholder-Employee Cooperative Game

R. F. LUCAS — Tariffs, Nontraded Goods, and the Optimal Stabilization Policy

D. O. PARSONS — Unemployment, the Allocation of Labor, and Optimal Government Intervention

M. FELDSTEIN — Fiscal Policies, Inflation, and Capital Formation

H. R. VARIAN — A Model of Sales

P. L. CHENG AND R. R. GRAUER — An Alternative Test of the Capital Asset Pricing Model

R. W. FOGEL AND S. L. ENGERMAN — Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Reply

D. L. BRITO AND W. H. OAKLAND — On the Monopolistic Provision of Excludable Public Goods

J. EATON AND H. S. ROSEN — Taxation, Human Capital, and Uncertainty

E. HELPMAN AND A. RAZIN — Efficient Protection under Uncertainty

SHORTER PAPERS: J. K. Stephens; R. Van Order; H. Yoshikawa; Y.-K. Ng; L. Samuelson; S. I. Cohen; M. K. Salemi; J. A. Frenkel; R. Driskill; W. M. Crain and A. Zardkoohi; D. N. Stevens and J. E. Foster; F. H. Stephen; E. G. Furubotn; J. S. Hekman; H. Herberg and M. C. Kemp; J. P. Neary; P. Howitt and D. Patinkin; R. Dusansky; I. M. McDonald.

SEPTEMBER 1980

THE AMERICAN ECONOMIC ASSOCIATION

- Published at George Banta Co., Inc. Menasha, Wisconsin. The publication number is ISSN 0002-8282.

- *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Effective April 1, 1980, Association membership dues for 1980, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$28.75 for regular members with rank of assistant professor or lower, or with annual income of $13,800 or less;

$34.50 for regular members with rank of associate professor, or with annual income of $13,800 to $23,000;

$40.25 for regular members with rank of full professor, or with annual income above $23,000;

$14.25 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $43.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $4.75 to cover extra postage.

- Correspondence relating to the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

- Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212.

Founded in 1885

Officers

President
MOSES ABRAMOVITZ
Stanford University

President-Elect
WILLIAM J. BAUMOL
Princeton University and New York University

Vice Presidents
CARL F. CHRIST
Johns Hopkins University
H. GREGG LEWIS
Duke University

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review
GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature
MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee
ROBERT W. CLOWER
University of California-Los Angeles
MARINA v. N. WHITMAN
General Motors Corporation
HENRY J. AARON
The Brookings Institution
ZVI GRILICHES
Harvard University
MARTIN S. FELDSTEIN
National Bureau of Economic Research, Inc. and Harvard University
ROBERT E. LUCAS, JR.
University of Chicago

Ex Officio Members
TJALLING KOOPMANS
Yale University
ROBERT M. SOLOW
Massachusetts Institute of Technology

MARGARET REID

DISTINGUISHED FELLOW
1980

Margaret Reid's long and distinguished career has been marked by complete integrity and dedication to the pursuit of verified knowledge. Her high standards and her single-minded devotion to her science have influenced generations of students and colleagues.

Margaret Reid was one of the pioneers in several areas of research on consumer and household behavior, each of which has now burgeoned into a major field of study on its own. For example, she did some of the earliest work on the concept and measurement of permanent income. Again, she was one of the first to see that one could systematically study the economics of the household use of time. And, of course, she has been a major contributor to the statistical analysis of the demand for housing. The empirical tradition at the University of Chicago owes much to Margaret Reid's example and teaching. She was a famous taskmaster in the art of applying critical thinking to data. Her reputation as a truly tireless researcher is well known and well deserved. Her warmth, her perceptive concern for students and colleagues, and her felicitous sense of humor are less widely known, but always cherished by those who worked with her.

Margaret G. Reid

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

ALBERT ANDO
PRANAB K. BARDHAN
PETER A. DIAMOND
W. ERWIN DIEWERT
RUDIGER DORNBUSCH
HERSCHEL I. GROSSMAN
PETER W. HOWITT
ANNE O. KRUEGER
WILLIAM H. OAKLAND
MICHAEL PARKIN
ROY RADNER
RICHARD W. ROLL
NANCY L. SCHWARTZ
JAMES P. SMITH
A. MICHAEL SPENCE
WILLIAM S. VICKREY
ROBERT D. WILLIG
S. Y. WU

●Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to George H. Borts, Managing Editor, Box Q, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

●No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

●Copyright American Economic Association 1980. All rights reserved.

September 1980

VOLUME 70, NUMBER 4

Articles

The Slowing Down of the Engine of Growth
W. Arthur Lewis 555

The Dynamics of Spot and Forward Prices in an Efficient Foreign Exchange Market
Jerome L. Stein 565

Experiments with a Decentralized Mechanism for Public Good Decisions *Vernon L. Smith* 584

A Model of the Firm as a Stockholder-Employee Cooperative Game *Masahiko Aoki* 600

Tariffs, Nontraded Goods, and the Optimal Stabilization Policy *R. F. Lucas* 611

Unemployment, the Allocation of Labor, and Optimal Government Intervention
Donald O. Parsons 626

Fiscal Policies, Inflation, and Capital Formation
Martin Feldstein 636

A Model of Sales *Hal R. Varian* 651

An Alternative Test of the Capital Asset Pricing Model
Pao L. Cheng and Robert R. Grauer 660

Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Reply
Robert W. Fogel and Stanley L. Engerman 672

On the Monopolistic Provision of Excludable Public Goods
Dagobert L. Brito and William H. Oakland 691

Taxation, Human Capital, and Uncertainty
Jonathan Eaton and Harvey S. Rosen 705

Efficient Protection under Uncertainty
Elhanan Helpman and Assaf Razin 716

Shorter Papers

Unemployment, Inflation, and Monetarism—A Further Analysis:
Comment *J. Kirker Stephens* 732
Reply *Robert Van Order* 738
On the "*q*" Theory of Investment *Hiroshi Yoshikawa* 739
Optimal Corrective Taxes or Subsidies when Revenue Raising Imposes an Excess Burden *Yew-Kwang Ng* 744
The Golden Rule with Endogenous Labor Participation Rate *Larry Samuelson* 752
Incentives and the Choice of Optimal Plans *Susan I. Cohen* 756
The Forward Exchange Rate, Expectations, and the Demand for Money—The German Hyperinflation:
Comment *Michael K. Salemi* 763
Reply *Jacob A. Frenkel* 771
Exchange Rate Dynamics, Portfolio Balance, and Relative Prices *Robert Driskill* 776
X-Inefficiency and Nonpecuniary Rewards in a Rent-Seeking Society: A Neglected Issue in the Property Rights Theory of the Firm *W. Mark Crain and Asghar Zardkoohi* 784
Comparing Utility Functions in Efficiency Terms:
Comment *Dana N. Stevens and James E. Foster* 793
Reply *Burton A. Weisbrod* 795
Bank Credit and the Labor-Managed Firm:
Comment *Frank H. Stephen* 796
Reply *Eirik G. Furubotn* 800
Income, Labor Supply, and Urban Residence *John S. Hekman* 805
In Defense of Some "Paradoxes" of Trade Theory *Horst Herberg and Murray C. Kemp* 812
This Side of Paradox, or, in Defense of the Correspondence Principle: A Reply to Herberg and Kemp *J. Peter Neary* 815
Utility Function Transformations and Money Illusion:
Comment *Peter Howitt and Don Patinkin* 819
Reply and Further Results *Richard Dusansky* 823
A Further Comment *Peter Howitt and Don Patinkin* 826
On the Comparison of the Stability Implications of Marshallian and Walrasian Adjustment Schemes: Note *Ian M. McDonald* 829

Notes 834

Editor's Note: The editorial office of the *American Economic Review* will move to the University of California-Los Angeles, January 1, 1981. The new managing editor will be Professor Robert W. Clower. Starting October 1, 1980, please address all new manuscripts to the *American Economic Review*, Editorial Office, University of California, Los Angeles, CA 90024.

Correspondence with regard to manuscripts submitted before October 1 should be sent to the present editorial office at Brown University, Providence, RI 02912.

The Slowing Down of the Engine of Growth

By W. ARTHUR LEWIS*

Let me begin by stating my problem. For the past hundred years the rate of growth of output in the developing world has depended on the rate of growth of output in the developed world. When the developed grow fast, the developing grow fast, and when the developed slow down, the developing slow down. Is this linkage inevitable? More specifically, the world has just gone through two decades of unprecedented growth, with world trade growing twice as fast as ever before, at about 8 percent per annum in real terms, compared with 0.9 percent between 1913 and 1939, and less than 4 percent per annum between 1873 and 1913. During these prosperous decades, the less developed countries (*LDC*s) have demonstrated their capacity to increase their total output at 6 percent per annum, and have indeed adopted 6 percent as the minimum average target for *LDC*s as a whole. But what is to happen if the more developed countries (*MDC*s) return to their former growth rates, and raise their trade at only 4 percent per annum: is it inevitable that the growth of the *LDC*s will also fall significantly below their target? My purpose is not to predict what is going to happen, but to explore existing relationships and how they may change.

The extraordinary growth rates of the two decades before 1973 surprised everybody. We knew that the world economy experiences long swings in activity; that world trade, for example, grew faster between 1830 and 1873 than it grew between 1873 and 1913, that is to say, between 4 and 5 percent before 1873, compared with between 3 and 4 percent after 1873. But a jump to 8 percent was inconceivable.

Some people were even more surprised by the performance of the *LDC*s. In 1950 these people were sceptical of the capacity of *LDC*s to grow rapidly because of inappropriate attitudes, institutions, or climates. The sun was thought to be too hot for hard work, or the people too spendthrift, the government too corrupt, the fertility rate too high, the religion too other worldly, and so on. This kind of analysis has now almost completely gone from the literature. In discussions at the end of the 1940's, noting that the *U.S.* national income per head had grown at about 2 percent a year over long periods, and noting that *LDC* populations were growing at about 1 percent, economists in United Nations circles were boldly discussing the possibility that the *LDC* growth rate might be 3 percent. The United Nations at the end of the 1950's fixed 5 percent as the target for the 1960's, meaning by "target" something that was unattainable but inspirational. Then to everybody's surprise UN figures showed 5 percent being averaged already by the middle 1960's. So the target was raised to 6 percent for the 1970's, but the figures showed 6 percent already in the early 1970's, and the UN was just getting ready to fix 7 percent for the 1980's when the recession started in 1974. I do not vouch for the accuracy of any of these performance figures, but I think *LDC*s have demonstrated beyond doubt their capacity to use physical and human resources productively.

The fast pace of world trade also played havoc with development theory. The collapse of international trade in the 1930's had seemed irreversible, so much so that Keynes had even declared that we didn't need much of it anyway. So in the 1940's and 1950's we created a whole set of theories which make sense if world trade is stagnant—balanced growth, regional integration, the two-gap model, structural inflation—but which have little relevance in a world where trade is growing at 8 percent per annum. Also many countries, basing their policies on the same assumption, oriented inwards mainly towards import substitution. The fact that world trade was growing rapidly was not universally recognized until the second half of the 1960's. Then nearly every country discovered the virtues of exporting. Now we are in danger of being caught out again. Since 1973 the growth rate of world trade has halved, and nobody knows whether this is temporary or permanent. But most of our economic writing continues to assume implicitly that a return to 8 percent is only just around the corner.

---

*Princeton University. This article is a revised version of the lecture W. A. Lewis delivered in Stockholm, Sweden, December, 1979, when he received the Nobel Prize in Economic Science. The article is copyright © the Nobel Foundation 1979. It is published here with the permission of the Nobel Foundation.

I

Let me come back to the relationship between *MDC*s and *LDC*s. The principal link through which the former control the growth rate of the latter is trade. As *MDC*s grow faster, the rate of growth of their imports accelerates and *LDC*s export more. We can measure this link. The growth rate of world trade in primary products over the period 1873 to 1913 was 0.87 times the growth rate of industrial production in the developed countries; and just about the same relationship, about 0.87, also ruled in the two decades to 1973.[1] World trade in primary products is a wider concept than exports from developing countries, but the two are sufficiently closely related for it to serve as a proxy. We need no elaborate statistical proof that trade depends on prosperity in the industrial countries. More interesting is the evidence that the relationship was quantitatively the same over a hundred years, so that the two-thirds increase in the rate of growth of exports of primary products from *LDC*s was no more or less than could be predicted from the increased rate of growth of *MDC* production.

Most interesting is that the coefficient is less than one, viz. 0.87. This means that if the engines of growth were industrial production in *MDC*s and exports of primary products in *LDC*s, then the *MDC* engine was beating slightly faster than the *LDC* engine. The effects of equal beating would not necessarily be exactly the same. And there are side effects that strengthen the connection. When the beat is faster the terms of trade are expected to be more favorable to the *LDC*s (though that did not happen this time). The domestic market prospers, so *LDC* industrialization for the domestic market is speeded; this happened. The *MDC*s relax their barriers to imports of manufactures, so this trade accelerates as well. Foreign capital flows into minerals, manufactures, and infrastructure. And foreign countries take more migrants, so that the homeward flow of remittances to *LDC*s is larger in prosperous times.

Putting it all together, including the fact that industrial production grew faster in *LDC*s than in *MDC*s, it is not surprising that the rate of growth of gross domestic product was just about the same in *LDC*s and in *MDC*s over the quarter of a century ending in 1973, namely about 5 percent per annum. Since *LDC* population was growing faster than *MDC* population, there is a big gap in the growth rates of output per head, about 4 percent in *MDC*s, against 2.5 percent in *LDC*s. The performance of *LDC*s was remarkable in absolute terms, but the gap between *MDC*s and *LDC*s in income per head continued to widen rapidly.

Now we come to our dilemma. The objective of most people who are concerned with these matters is to narrow the per capita gap between *MDC*s and *LDC*s. But how is one to do this if they are linked to equal growth of total output? One might perhaps conceive a lower rate of growth of *MDC*s. Many *MDC* voices are calling for this—the environmentalists, the persons who fear exhaustion of exhaustible resources, the advocates of greater grace and leisure in our lives, and others. But if the *MDC* growth rate falls, the *LDC* growth rate will fall too, and *LDC*s will get the worst of it, since the terms of trade will move against them. Given the link, it is in the interest of *LDC*s that the *MDC*s should grow as rapidly as they can.

---

[1] For data, see my *Growth and Fluctuations 1870–1913*, pp. 175–76.

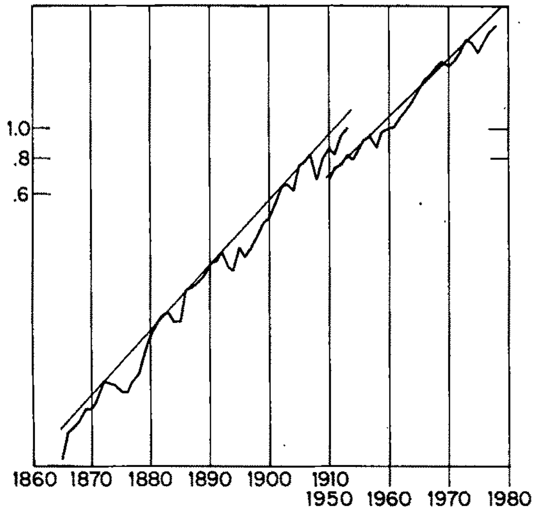


FIGURE 1. U.S. INDUSTRIAL PRODUCTION, 1865-1913 AND 1950-78

Three questions come to mind. First, what is likely to happen to the speed of *MDC* growth? Secondly, can *LDC*s maintain fast growth even in face of a decline of *MDC* growth? And thirdly, is it desirable that *LDC* development depend on rapid growth of exports to *MDC*s? My main purpose in this paper is to consider the second question, but I shall also briefly explore the first. The third question, on the desirability of *LDC/MDC* trade, may be left to another occasion.

II

Many people now assert that the world economy has made one of those major turns that it makes from time to time, as it did in 1873 when world trade settled into a growth rate only about two-thirds of that of the preceding half-century. However, part of the evidence they adduce is merely evidence of cyclical and not secular decline—high unemployment rates, low profits,[2] low investment ratios, low savings, and a slower growth rate of productivity are the familiar elements of cyclical downswing, and throw no light-term trends.

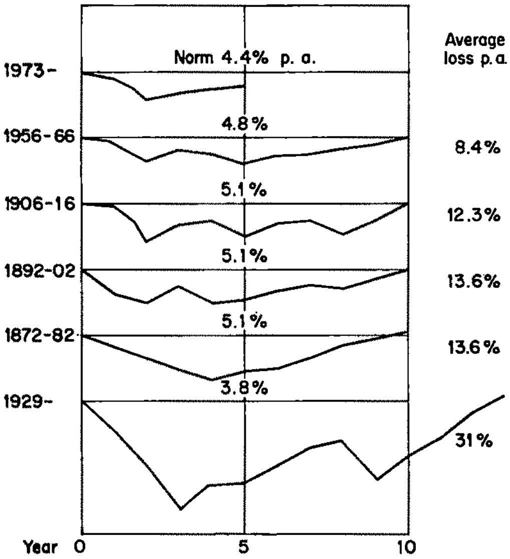


FIGURE 2. U.S. GREAT DEPRESSIONS: DEVIATION OF ACTUAL FROM "CAPACITY" INDUSTRIAL PRODUCTION

Over the past century, the United States has experienced a series of Great Depressions, each of which took ten years to complete itself, except for that of 1929, which took twelve years. The starting points of these depressions were 1873, 1893, 1907, 1929, 1957, and 1974. Figure 1 shows *U.S.* industrial production on a semilogarithmic scale, where the straight lines indicate the rate of growth along the peaks, and the potential output at that rate.[3] The difference

---

[2] Evidence of a decline in the profits ratio, often cited, is hard to interpret. In the United States, industrial production runs along a ceiling when the ratio of manufacturers' profits to sales is around 5 percent (after tax). The ratio averaged 4.8 over 1951–56 and 5.1 over 1966–69, which were also years of full employment (less than 4 percent unemployed). The basis of the series changes between 1973 and 1974; on the new basis, the ceiling would be about 6 percent. That the profit ratio averaged 5.4 in 1976 and 5.3 in 1977 merely indicates, along with elevated unemployment, the continuance of cyclical deficiency in those years. Series from the *Economic Report of the President*.

[3] From 1923, the index is that of the Federal Reserve Board. Prior to 1913, the index is that of Edwin Frickey, as amended by the National Bureau of Economic Research statisticians. For references, see my *Growth and Fluctuations 1890-1913*, pp. 272-73. Construction is excluded. The charts are semilogarithmetic. The slope of the straight line from 1872 to 1906 indicates growth at 5.1 percent per annum; that from 1951 to 1973

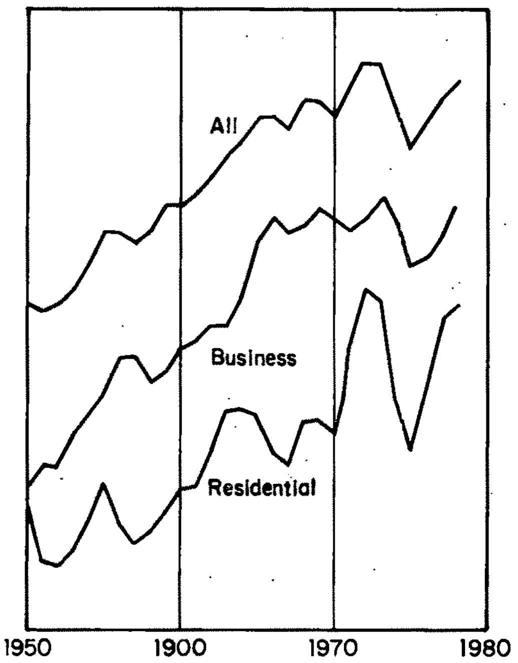


FIGURE 3. U.S. CONSTRUCTION

between actual and potential output is shown in Figure 2 for each of these Great Depressions separately.[4] The depression at the top of the figure is the one we are now experiencing. It seems to conform to pattern. For example, as is well known, these Great Depressions are associated with similar long fluctuations in construction activity. Figure 3 shows that this association is unchanged, despite Moses Abramovitz's optimistic prediction to the contrary; deep depressions in both business and residential construction are at the heart of our current depression.[5] Again, suggestions that the *U.S* economy was overheated at the end of 1978 (year 5) seem implausible—they confuse lack of capacity due to many years of low construction in face of rising population, with lack of capacity due to high demand. Shortage of capacity after years of depression is what triggers a new investment boom. If the past repeats itself, a spurt of activity in the early 1980's (growth at 6 to 7 percent per annum) should bring *U.S.* industrial production back on to its prosperity line by 1983. This does not rule out the possibility of a minor fluctuation in the interim. There is, of course, no basis for predicting that the past will repeat itself, but equally there is no basis for predicting that this depression will go on forever.[6]

If then we reject all evidence that can be explained by six years of depression, we are left with a number of arguments that are being advanced to suggest that the prosperity of 1950–73 was special and not repeatable. I shall merely list them, because to pursue each of them would take us too far off course. Here are the leading six:

1. The fast growth of Europe after World War II was due to catching up on a backlog of innovations whose feasibility and profitability the United States had demonstrated by 1950, but whose utilization had been delayed in Europe by two world wars and the Great Depression — telephones, automobiles, refrigerators, television, aeroplanes, and so on. This backlog, it is claimed, is now exhausted.

2. There is no new innovation of Schumpeterian magnitude to take its place. This proposition could be established only by hindsight; expenditure on research and development by private industry may have declined, but this may only be a cyclical event.

3. Western Europe's reserves of surplus labor in agriculture, petty retailing, and

---

indicates growth at 4.4 percent per annum. In Figure 1, the position of the lower curve in relation to the upper reflects that industrial output in 1956 was 20 percent less than it would have been if it had grown at 4.4 percent per annum since 1913. The interwar curve is omitted. Its familiar shape would show output rising at 3.8 percent per annum before 1929, then plunging into the Great Depression.

[4] The depression of 1929 is of such a different magnitude from the rest that it has been taken out of its sequence, and put at the bottom of the figure.

[5] Expenditure on structures at 1972 prices; Tables B2 and B7 of the *Economic Report of the President*. It can be seen in Figure 2 that the two postwar depressions are notably less deep than the pre-1913 depressions.

[6] The National Bureau of Economic Research has trained us to think of the Kitchin as the reference cycle (forty months), so any depression that extends beyond eighteen months feels like the end of the world. Our economic journalists have even lower horizons, since they seldom get beyond comparing this year with last. We cannot understand what is happening without at least a ten-year perspective.

elsewhere, which facilitated rapid expansion of industry and other high-level occupations is now exhausted, and immigration of cheap labor from Southern Europe will not be resumed.

4. We shall run into a shortage of minerals; of copper, tin, bauxite, and others, because transnational companies and *LDC* governments cannot agree on terms for new investment; and of oil because of OPEC's conservation policies. This imbalance is one of the two primary sources of inflation. The other primary source is the spiralling of prices as organized groups bargain for an annual increase in income of at least 5 percent (plus last year's increase in prices) in an economy where productivity rises by 3 percent per annum or less. Attempts to control inflation through controlling the money supply touch neither of these primary sources, unless pushed to the level of 15 percent unemployment, which is not politically feasible. Whether the monetary authorities opt for stagflation or for massive unemployment, the rate of growth is slowed.

5. The preference of consumers in rich countries for services rather than manufactured commodities will result in relative decline of the industrial population. Among the effects of this may be a decline in the rate of growth of imports of primary production.

6. High levels of taxation may diminish initiative, enterprise, and the rate of growth.

I list these propositions neither to support them nor to controvert them, but only to remind us that the world economy has had long periods of prosperity (like 1850 to 1873) and long periods of relative stagnation (say 1913 to 1950) so that there is nothing strange in the idea that the next two or three decades may turn out to be difficult. But there is also nothing strange in the idea that they may turn out to be rather prosperous.

III

In what follows I shall assume that industrial production in *MDC*s grows more slowly than it was growing before 1973, and that the imports of these countries grow only at 4 percent a year, over the next twenty years. This is not a prediction; it is merely the assumption whose consequences we are seeking to analyze.

I shall also assume that *LDC*s want their *GDP* to grow at 6 percent per annum, and that this requires their imports to grow at 6 percent. This linkage follows from the further assumption that the individual *LDC* will not become more self-sufficient, perhaps because it is too small; though *LDC*s as a group will have to be more self-sufficient. No importance attaches to whether the figure for imports is the same as the figure for the growth of gross domestic product; all that matters is that the growth rate of exports from *LDC*s is assumed to be significantly higher than the rate of growth of imports of *LDC* commodities into *MDC*s. The *LDC*s will continue to pay for some of their imports out of proceeds of transfers, including foreign aid and private foreign investment, but we shall assume that this still leaves *LDC*s needing, say, a 6 percent growth rate for exports, while *MDC*s are assumed to increase their imports from *LDC*s only at 4 percent a year. The problem is how to reconcile these two growth rates.

There could theoretically be a simple way, namely for *LDC*s to have an ever increasing share of *MDC* imports, but we have closed this door. The main link between *MDC* and *LDC* economies has been the *MDC* demand for *LDC* primary commodities. This has been a link in terms of physical volume, not much affected by prices. The *LDC*s could not sell significantly more primary produce by reducing prices; on the contrary, they would earn substantially less purchasing power as the terms of trade deteriorated. The *LDC*s could earn more by reducing the volume or by joining together in raising prices. The direct effect of these actions would be to reduce output, but this could be offset by investing the extra earnings judiciously. However none of this seems to be in the cards; so we shall assume that our problem cannot be solved by accelerated or decelerated production of primary commodities normally exported to *MDC*s.

What about manufactures? These are now nearly 40 percent of the exports of the non-OPEC *LDC*s, and are still their fastest growing export. Could the whole problem be solved simply by increasing the growth rate of manufactured exports to *MDC*s, in substitution for primary products? I shall assume that this cannot be done, since if it can be done my paper ends abruptly. Also I do not think that it can be done. The *MDC*s are willing to let in manufactured exports when they are prosperous, since they then have many growing industries that can take in people displaced by imports. Our assumption that the *MDC* growth rate is low rules out this possibility. It would indeed be more appropriate to assume that *MDC*s will take less manufactures from *LDC*s rather than more.

Our basic assumptions therefore are that *LDC*s need to have their exports grow at 6 percent a year, but *MDC*s will increase their imports from *LDC*s only at 4 percent a year. What is to happen to the growth of *LDC* output?

Let me concede at once that from the standpoint of the individual *LDC* it matters not at all what the *MDC* growth rate may be. Given resources and flexibility, it can always sell more to *MDC*s. However, it thereby displaces some other *LDC*'s trade. What one can do cannot be done by all.

At the level of arithmetic this problem now has only one solution. If total sales from *LDC*s increase at 6 percent, while sales to *MDC*s increase at 4 percent, sales to the rest of the world (given weights of seven to three) must increase initially at about 11 percent per annum. Ignoring the socialist countries, which could help by buying much more from *LDC*s but won't, the *LDC*s can solve the problem only by accelerating sharply their trade with each other.

Inter-*LDC* trade is still rather small—about 19 percent of the exports of non-OPEC *LDC*s.[7] The percentage did not change significantly over the two decades to 1973 despite all the effort that was put into creating and servicing regional trade institutions. Can this trade take up the slack left by *MDC*s as *MDC*s slow down?

The answer is in the affirmative. Currently the *LDC*s depend on the *MDC*s for food, fertilizers, cement, steel, and machinery. Taken as a group, *LDC*s could quickly end their dependence for the first four, and gradually throw off their dependence for machinery. They also import a considerable quantity of light manufactures for which they are not in any sense dependent (some $31 billion in 1977, compared with $47 billion of engineering products). They could quickly rid themselves of these, and more gradually throw off their dependence for machinery.

The *LDC*s are capable of feeding themselves now, if they adopt appropriate agrarian policies and, as our eleven new international tropical agricultural research institutes give us better varieties and improved technology, output should more than keep up with population. The problem is to get through the period while the birth rate remains obstinately high to the less-frightening times when the birth rate will have dropped below twenty per thousand. It may be a near thing, but we should make it.

As for fertilizers, cement, and steel, these are made by applying standard technology to raw materials that are widely available outside the *MDC*s. Machinery is more bothersome because important parts of this trade involve economies of scale, continually improving technology, and patented or secret knowledge. However, several *LDC*s are moving into this field, and already machinery is 15 percent or more of the output of manufactures in at least eight *LDC*s (India, Brazil, Singapore, Chile, Korea, Argentina, Mexico, and Israel).[8] The *LDC* exports of engineering products are also growing rapidly, and contrary to popular belief, already exceed *LDC* exports of textiles and clothing in value. There is no reason why *LDC*s as a group should not become nearly self-sufficient in standard types of equipment.

If all this scope exists for inter-*LDC* trade, why have the regional customs unions not been more successful? Note three reasons.

First, a region is not a homogenous area. Some countries are much more advanced than others in industrial competitiveness, to an almost inconceivable extent. These advanced countries attract more new industries than the less advanced, who feel exploited by the customs union. The union then survives only if costly measures are taken to placate the less advanced, and these measures are difficult to negotiate.

---

[7] Trade data in this and the next two paragraphs are from GATT, *International Trade 1977–78.*

[8] Share of machinery in manufactures calculated from United Nations, *Yearbook of Industrial Statistics 1976.*

Secondly, the usefulness of the union is maximized in sharing out industries with substantial economies of scale, extending over the whole regional market. Each country is anxious to keep for itself all those industries that can attain the economies of scale within the national market. Destruction of any of these industries by competition from another member of the union causes a political uproar. The union is therefore safest when it does not require internal free trade in all commodities, but instead concentrates on those few "integration" industries as they are called which need the whole regional market. Even this more modest task is hard to negotiate if each member country is to have its fair share of integration industries.

The third reason why the customs unions have not done better is that their basic assumption—that a country should trade most with its next-door neighbors—is no longer true in these days of very low transport costs. For reasons of climate, soil, and history, the next-door neighbor is probably in the same business, and not a potential customer or supplier. He is equally poor, and therefore offers an equally limited market. The *LDC*s developing new industrial products are drawn to large rich *LDC* markets rather than to those of their neighbors. Frankly, in the 1950's and 1960's, aggressive developers did not much need the support of customs unions; these offer more to their members when world trade is stagnant than when it is booming.

It follows, therefore, that in the situation we are analyzing, where world trade decelerates, customs unions would be more highly prized and would be made more effective, especially in regard to large-scale industries with region-wide economies of scale. But even so, the leading commodities that *LDC*s would now have to produce to a greater extent for each other cannot be shared out between next door neighbors on a political basis. Food, fertilizers, cement, steel, and fuel pick their locations more in terms of raw material availability, and machinery will come in the first instance from those *LDC*s that already have a substantial industrial base. This new *LDC* trade would be worldwide, just as European and *U.S.* trade are worldwide.

IV

I am therefore arguing that it is physically feasible for *LDC*s to maintain a high rate of growth even if *MDC*s decide otherwise for themselves. How does physical feasibility translate into an effective economic framework?

One way would be to follow the customs union route, with *LDC*s giving preferential treatment to imports from other *LDC*s. The nucleus of this exists already in the Protocol Relating to Trade Negotiations among Developing Countries which came into force in 1973, with the blessing of GATT, and which provides for negotiated preferential arrangements among sixteen of the bigger and more advanced *LDC*s. The philosophy of such an arrangement is in line with the spirit of Bretton Woods, which recognized the rights of countries to impose restrictions on other countries which were tending persistently to run balance-of-payments surpluses, as would be the situation if the *LDC*s were growing faster than the *MDC*s. One may doubt however whether such different countries will get very far along the route of preferential concessions. If they are to prefer each other's goods, this will have to be because they are competitive in price with those of *MDC*s.

In the economist's model this competitiveness would come about automatically. The *LDC*s would run a balance-of-payments deficit because of the *MDC* slowdown, yet persist in their own rapid growth instead of slowing down themselves. Adjustment comes in the old gold standard version through an outflow of gold that reduces their price levels; or in the modern versions by devaluation, which has the same effect. The real world is more complicated. Inflation is universal, but aggressive sellers of manufactures have to keep their prices down; so this set of *LDC*s would need special emphasis on inflation controls. Devaluation cannot be avoided when prices cease to be competitive, but is palliative rather than curative in situations when it triggers further increases in domestic costs that reinstate the differential that it was meant to eliminate. The *LDC*s have the same problem as the *MDC*s: the domestic price level can no longer be controlled merely by twirling general controls such as the rate of interest, or the supply of money, or the rate of exchange for the currency. They too now experience the cost-push element in price determination, for which the only remedy is some sort of incomes policy. We expect more from the economic system than our grandparents did in the nineteenth century, by way of full employment and faster growth, and should not be surprised that the economic system requires more from us, by way of supporting institutions.

These new aggressive *LDC*s, exporting machinery to each other, may also have problems in financing their trade. Nearly every *LDC* has a separate currency. We are envisioning Nigeria selling cereals to India for rupees, with which it buys machinery from Brazil. Some kind of clearing agreement may become necessary; otherwise *LDC* traders will tend to do business with each other in one or more *MDC* currencies, and will be constrained by the relative scarcity of such currencies. Perhaps the International Monetary Fund would straighten this out. A more serious problem will be to finance the export of capital goods from one *LDC* to another, since the seller is expected to help finance the buyer. It is not likely that the *LDC* exporters can do this on their own. We must assume that they will be allowed to raise untied loans in the *MDC* financial markets, perhaps using the regional development banks as intermediaries to a greater extent than is now the case.

But the real problem is not whether *LDC*s can become competitive and hold their own in each other's markets. Problems of pricing and foreign exchange can work themselves out in the world market. The real problem is whether *LDC*s will persist in rapid growth despite the slowdown of the *MDC*s. If the economy is still dependent, the balance-of-payments weakness will pull it down; but if it has attained self-sustaining growth, the weakness in the foreign exchanges merely launches a drive to export to other *LDC*s, and the weakness in the balance of payments is then only transitional.

If a sufficient number of *LDC*s reach self-sustaining growth, we are into a new world. For this will mean that instead of trade determining the rate of growth of *LDC* production, it will be the growth of *LDC* production that determines *LDC* trade, and internal forces that will determine the rate of growth of production. Not many countries are ready to make this switch. India is an obvious possibility, along with some of the other subscribers to the Protocol of 1973. It is not possible for all *LDC*s to make this switch and neither is it necessary, for if leading *LDC*s grow fast and import heavily, they will substitute to some extent for the former rapid growth of *MDC*s. For those who use the language of center and periphery, this means that a number of countries leave the periphery and join the center. Or if they are specially linked to each other by preferential trade and currency arrangements, one may even speak of the creation of a new center consisting of former peripheral nations that have built a new engine of growth together.

The shadow on this picture is what happens to those *LDC*s whose best option has been to export raw materials to *MDC*s. Our exercise starts from the assumption that the growth rate of *MDC* demand is reduced, so these face surpluses and unfavorable terms of trade. We have provided an escape for *LDC*s that can turn to exporting food or manufactures, but we have not assumed that the new core *LDC*s will substitute for the *MDC*s by drinking more coffee or tea, or using more rubber and jute. This solution therefore involves some hardships for the less adaptable *LDC*s, constrained by climate or by the small size of their markets. A framework for helping them exists already in the IMF's compensatory financing, and in the EEC's STABEX support; but these are meant for temporary fluctuations. Bigger and more persistent support would be required.

Transnational corporations would probably play some part in the establishment of this new inter-*LDC* trading network. The cadre of domestic entrepreneurs is adequate in the more advanced industrial *LDC*s to manage most of the range of light consumer goods and light engineering products. One of our main concerns, however, is to diminish reliance on *MDC*s for heavy machinery and such, and this extends into fields where experience is limited. Since we are assuming that the market will send out price signals favoring production in *LDC*s, whether because of tariffs or of currency adjustments, transnational corporations will be eager to preserve their markets by establishing subsidiaries behind the protective barrier. Hostility to such corporations is universal and their influence is diminishing in most sectors, especially in mining, public utilities, distribution, and finance; but not in manufacturing, where judging by advertisements in the *New York Times* and the financial press, *LDC* governments are only too anxious to invite the participation of transnational corporations. There are plenty of restrictions—on the hiring of expatriate staff, on percentages of equity owned by foreigners, on borrowing in the local market, on technology, and so on. Also, in many cases joint ownership with local capitalists or with government agencies is prescribed. A government anxious to promote industrialization and a corporation anxious to preserve or extend its market find common ground.

The awkward part of the exercise is to sustain the momentum of 6 percent growth through the transition from dependence on *MDC* trade to dependence on *LDC* markets. During this transition the leading industrial *LDC*s must establish their footholds in each others' markets, as well as those of other *LDC*s; also the agrarian changes must occur which both feed the urban population and present a growing market for its goods and services. It is possible that some of the leading *LDC*s can take this in their stride, just as German industrialists launched their trade drive in the 1880's, followed by the United States after 1895, by Japan in the 1930's, and more recently by Brazil. They do not have to begin with machinery, since *LDC*s still import so much light manufactures from *MDC*s. They can start here and move more gradually into machinery.

At the other extreme, it is also possible that there is simply not yet enough entrepreneurial steam in the leading *LDC*s to make this transition without a supporting framework. I have already mentioned the main international elements of such a framework, namely preferential tariff and currency arrangements. The domestic element consists of the maintenance of home demand in the face of stagnant world trade in primary products, so that the economy can continue to go forward instead of collapsing. Much of the responsibility for maintaining momentum then falls on the government, given its large share of the cash economy, and also the extent to which it regulates or supports the private sector. It has to carry the responsibility for a large investment program (private and public) in human and physical capital. This responsibility could not be carried without external aid. The *MDC*s would have to be in a mood to say: we will not give you more trade; here for a while is more aid instead.

The recession that started in 1974 has now lasted long enough for *LDC*s to consider the possibility that *MDC*s intend to maintain rates of *GNP* growth which will allow world trade to expand only at around 4 percent a year. This would be a major blow to *LDC* growth aspirations, unless new steps were taken to support rising participation of *LDC*s in *LDC* trade. I have been trying to analyze what these steps might be. They ought to figure prominently in current North-South negotiations, but in fact they do not, since these negotiations tacitly assume that high *MDC* growth rates will soon be resumed. Perhaps this assumption will turn out to be correct; economics does not foretell the future. For the least, *LDC*s should be discussing among themselves in what directions they would wish to go, prior to negotiations with the *MDC*s.

Of course, the problems tackled in this paper would not arise at all if *MDC*s were willing to allow *LDC*s a greater share of *MDC* markets. This would be the logical evolution of a situation where *LDC*s grow faster than *MDC*s; trade with *LDC*s should become an ever-increasing portion of *MDC* trade. We live in a strange world. Through the 1960's and 1970's, *MDC*s have been dismantling their barriers to each other's trade while increasing their barriers to *LDC* trade. Since imports of manufactures from developing countries are only 2 percent of the consumption of manufactures in OECD countries, this indicates exceptional sensitivity to minor change. Lack of sensitivity, on the other hand, characterizes the failure of developed countries to recognize that dependence is mutual, in that the non-OPEC *LDC*s take 20 percent of OECD exports, and could therefore by their prosperity help a little to sustain OECD prosperity. It can hardly be an OECD interest to force the *LDC*s into discriminating against OECD sources.

Nor would these problems arise if the *MDC*s would return to the attack on mass poverty within their own borders which they launched so successfully in the 1950's and 1960's, and have now abandoned; since what we all really need is that world trade recapture its growth rate of 8 percent a year. But that is a different story.

REFERENCES

M. Abramovitz, "The Passing of the Kuznets Cycle," *Economica*, Nov. 1968, *35*, 349–67.

W. Arthur Lewis, *Growth and Fluctuations 1870–1913*, London; Boston 1978.

GATT, *International Trade 1977–78*, Geneva 1978.

United Nations, *Yearbook of Industrial Statistics 1976*, Vol. 1, New York 1978.

UNCTAD, *Handbook of International Trade and Development Statistics, Supplement 1977*, New York 1978.

U.S. Council of Economic Advisers, *Economic Report of the President*, Washington, various years.

The Dynamics of Spot and Forward Prices in an Efficient Foreign Exchange Market with Rational Expectations

By JEROME L. STEIN*

Many papers have examined the "efficiency" of the foreign exchange markets during the current float and during the previous adjustable peg period.[1] An "efficient" market is one in which the current price fully reflects all relevant available information concerning the evolution of the system from the conditions prevailing at time t to those that will occur at t+1. The research strategy underlying this set of papers is as follows: If the markets can be shown to be efficient, then the market is setting current rates in a rational manner, given the publicly available information. The "unduly large" fluctuations in the exchange rate[2] do not result from the activities of irrational speculators, but result either from the large random shocks that impinge upon the economy or from the erratic behavior of the goverments in affecting the rate of exchange.

The types of tests used to evaluate the efficiency of the foreign exchange market have been carried over from the stock market literature; and their results are ambiguous and confusing. Kohlhagen summed up the state of recent research:

> First...it seems that the basis for this literature—namely, what is conclusive evidence of an efficient market—has not been very well developed. Or at least it has not been universally accepted, as can be seen in the contradictory evidence presented in many of the studies. Second the techniques that are used suffer from being ad hoc applications from the stock market literature. That is, nobody has developed a notion of what an efficient foreign exchange market is and how it could be empirically tested, as a separate and distinct concept from what an efficient stock market...is....[p. 34]

It is impossible to evaluate whether any asset market is efficient in processing information without specifying a model which links prices expected to prevail at time t+1 to prices which are then determined in period t. A model which may be relevant for the stock market (for example, current prices are set to yield positive or constant expected returns) cannot be presumed to be relevant for the foreign exchange market. The aim of this paper is to describe the intertemporal pattern of spot and forward exchange rates in a dynamic model with rational expectations; that is, where anticipations are informed predictions of future events based upon the correct model.

My main conclusion is that the standard "efficient market" tests are not applicable to the foreign exchange (or any other) market where there is a feedback from the price, which equates the stock demand to the stock in existence, to the rate of change in the stock. The standard efficient market hypothesis tests whether the sequence of changes in the forward exchange rate, or of spot exchange rates adjusted for interest rate differentials, is a "fair game";[3] that is, has a zero expectation and is not serially correlated. It is proved here that:

---

*Brown University. Research on this paper began in 1976 when I was a visiting scholar at the Federal Reserve Board. I am indebted to George Borts, Marvin Goodfriend, Steven Kohlhagen, and Keehwan Park for valuable criticisms of earlier drafts, and to the participants at my seminars at the Federal Reserve Board (April 1978), the Hebrew University (May 1978), the Swiss National Bank (June 1978), and Princeton University (May 1979) for valuable suggestions.

[1] See the comprehensive description of these studies in Steven Kohlhagen.

[2] Michael Dooley and Peter Isard examined the dollar-deutschemark exchange rate from May 1973 to June 1977. They concluded that during the study period, the absolute values of month-to-month changes in the spot rate exceeded two cents per mark (an annual rate of 60 percent) for 6 out of 60 months, and exceeded 1 cent per mark during 20 months. Interest differentials were small relative to these exchange rate movements.

In a market with rational expectations, (i) if there exist both risk aversion and responsiveness of the basic balance of payments to the spot price, or (ii) if the "flow equilibrium price" which equates the basic balance to zero is changing, *then* neither the change in the spot price adjusted for interest rate differentials nor in the forward price will be a fair game. An examination of price sequences, per se, is incapable of determining whether the exchange market is efficient in processing information.

The basic balance of payments (B(t)) over any interval of time is defined as the sum of net sales of goods, services, and long-term financial instruments to foreigners, by residents of a country. In a free exchange market, the balance of payments (*BOP*) is zero.

BOP(t)$\equiv$(net sales of goods and services + net sales of long-term financial instruments) − (net purchases of short-term financial instruments)$\equiv B(\mathrm{t})-\Delta S(\mathrm{t})=0$.

The first term in parentheses is the basic balance B(t), the balance on current account plus the balance on long-term capital account. The second term is the short-term capital outflow ΔS(t), the change in net short-term claims against foreigners by private residents of a country. The economic question is: What mechanism of adjustment produces a zero *BOP* over every interval of time in a free exchange market when expectations are rational?

The accumulated sum of the basic balances $\Sigma_{\tau=0}^{t}B(\tau)$ is the stock of net short-term claims against foreigners, denoted by $S(\mathrm{t}) \gtreqless 0$, since the basic balance B(t) is equal to the change in the stock of net short-term claims against foreigners ΔS(t). Using portfolio theory, a stock demand function for S(t) is derived relating the (logarithm of the) price of foreign exchange p(t) to the stock S(t). The steepness of this curve depends positively upon risk aversion and negatively upon the sensitivity of the basic balance to the price of foreign exchange.

I define an equilibrium stock denoted by $S_e \gtreqless 0$. At any moment of time, the price of foreign exchange p(t) adjusts to produce an expected rate of return on the stock of net short-term claims against foreigners which induces the public to hold in their portfolios the stock S(t) in existence. At that price, the short-term capital flow ΔS(t) exactly offsets the basic balance B(t).

Basic balance B(t) is viewed as a function of the difference between the current price of foreign exchange p(t) and a value y(t) which is associated with a zero basic balance. This value y(t) could be the price of foreign exchange derived, for example, by the monetary approach to the *BOP* such that: (i) the excess demand for money is zero and (ii) relative purchasing power parity prevails.

The main result of our dynamic model with rational expectations is described schematically as follows.

$$\begin{tikzcd} \lbrack S(\mathrm{t})-S_e\rbrack \arrow[r, "(-)"'] & \lbrack p(\mathrm{t})-y(\mathrm{t})\rbrack \arrow[r, "(+)"'] & B(\mathrm{t})=\Delta S(\mathrm{t}) \arrow[ll, bend left, "\textit{Feedback}"] \end{tikzcd}$$

At any time, the "excess" stock of net short-term claims against foreigners $[S(\mathrm{t})-S_e]$ is predetermined. The price of foreign exchange p(t) adjusts to induce the public to willingly hold the existing quantity in their portfolios. This relation is based upon the stock demand function, and the exchange market clears. The resulting deviation $[p(\mathrm{t})-y(\mathrm{t})]$ of the price of foreign exchange from its longer-run equilibrium value affects the basic balance B(t). An excess stock of net short-term claims $S(\mathrm{t})-S_e>0$ leads to a price below the longer-run equilibrium $[p(\mathrm{t})-y(\mathrm{t})]<0$ and then to a negative basic balance, that is, to an excess flow demand for goods and long-term financial instruments. A negative basic balance means that ΔS(t) is negative: and the stock of net short-term claims against foreigners converges to the equilibrium value S_e. By the same process, the price of foreign exchange converges to

---

[3] See Appendix A below for a discussion of these concepts and a brief summary of empirical results.

TABLE 1—RETURNS OR COSTS ASSOCIATED WITH SHORT-TERM ASSETS OR LIABILITIES IN DIFFERENT CURRENCIES, HEDGED OR UNHEDGED

	Dollar Return or Cost	Expected Relative Rate of Return or Cost
Dollars	$P(t)e^{r_H}$	
Foreign currency:		
Hedged	$Q_{t+1}(t)e^{r_F}$	$x_1(t) \equiv [q_{t+1}(t) - p(t)] - [r_H - r_F]$ $\equiv x_2(t) - x_3(t)$
Unhedged	$P^*(t+1)e^{r_F}$	$U(t) \equiv [Ep^*(t+1;t) - p(t)] - [r_H - r_F]$ $\equiv [Ep^*(t+1;t) - p(t) - x_3(t)]$

Note: $p(t)$ is the logarithm of the dollar price of foreign exchange $P(t)$; $Ep^*(t+1;t)$ is the logarithm of the anticipated dollar price of foreign exchange at t+1, $P^*(t+1;t)$; $q_{t+1}(t)$ is the logarithm of the price of forward exchange $Q_{t+1}(t)$ for delivery at t+1; $x_2(t) \equiv q_{t+1}(t) - p(t)$ is the forward rate; $x_3(t) \equiv r_H(t) - r_F(t)$ is the domestic less foreign interest rate.

its equilibrium value $y(t)$, for example, as determined by the monetary approach to the balance of payments.

The speed of convergence depends upon: (i) the degree of risk aversion displayed by those who hold short-term claims against, or who have short-term liabilities to, foreigners, and (ii) the responsiveness of the basic balance to the deviation of $p(t)$ from $y(t)$.

From the dynamic analysis sketched above, the trajectories of the spot and forward prices are derived. Insofar as the price of foreign exchange $p(t)$ converges asymptotically to $y(t)$, the change in the spot price (adjusted for interest rate differentials) is not a fair game. It is shown that when there is feedback from $p(t)$ to the basic balance, there is no reason why the standard tests of market efficiency are relevant when expectations are rational.

I. The Basic Model

A. *Portfolio Choice*

Consider a world with two currencies: the dollar and the foreign currency (for example, yen), with a free rate of exchange, $P(t)$ dollars per yen. At any time there is a stock of net short-term claims against foreigners, *U.S.* claims against foreigners less foreign net claims against Americans, $S(t)$ which is the sum of the *U.S.* basic balances. Some of these claims are denominated in dollars and some in foreign currency, and they are appropriately converted into a common currency for summation. On the basis of the theory of portfolio choice, a demand for a stock of net short-term claims against foreigners is constructed.

A *U.S.* resident can hold short-term dollar assets or short-term foreign currency assets; and the latter can be held unhedged or hedged with a forward sale of foreign exchange. The portfolio choice is based upon the following relative rates of return.

Let the dollar price of a unit of foreign exchange at time t be $P(t)$.

(i) If $P(t)$ is invested in a dollar asset yielding a rate of return r_H per period, then $P(t)e^{r_H}$ dollars will be received at the end of one period. Alternatively, if the unit of foreign exchange is invested at the foreign interest rate r_F per period, then e^{r_F} units of foreign exchange will be received at the end of the period. There are two ways to convert e^{r_F} units of foreign exchange into dollars.

(ii) The owner of the asset denominated in foreign currency could wait until time t+1 when the asset matures and sell each unit of foreign exchange for $P^*(t+1)$ dollars. Thereby $P^*(t+1)$ is a stochastic variable (denoted by an asterisk).

(iii) Alternatively, at time t when the foreign exchange is initially purchased, it can be simultaneously sold forward for delivery at time t+1 at price $Q_{t+1}(t)$. Then $Q_{t+1}(t)e^{r_F}$ dollars will be received if the asset is held to maturity. The assumption of offsetting positions in the spot and forward markets is called *hedging*.

Column 1 of Table 1 lists the dollars to be received at time t+1 from these three alternatives. Column 2 contains the expected rate of return from hedged and unhedged foreign currency assets, relative to the return on domestic assets. Let lowercase letters denote the logarithms of corresponding uppercase letters, for example, $p(\mathrm{t}) \equiv lnP(\mathrm{t})$. The term $x_2(\mathrm{t}) \equiv q_{t+l}(\mathrm{t}) - p(\mathrm{t})$ is called the forward rate, $x_3(\mathrm{t}) \equiv r_H(\mathrm{t}) - r_F(\mathrm{t})$ is called the interest parity, $x_1(\mathrm{t}) \equiv x_2(\mathrm{t}) - x_3(\mathrm{t})$ is the net covered yield, and $U(\mathrm{t})$ is the expected yield on an uncovered position. The expectations or average operator is denoted by E.

Institutions may incur *liabilities* in either the domestic currency or the foreign currency. For example, a multinational firm may borrow either dollars or foreign currency to finance its operations; and it is obligated to repay in the currency initially borrowed. In this case, the firm is holding a negative asset. The cost of the liability (negative asset) is the same as the return on the corresponding asset. Consequently, column 2 of Table 1 also describes the expected costs of hedged and unhedged liabilities in foreign currency relative to the cost of a domestic currency liability.

If the domestic currency is the safe asset and the foreign currency is the risky asset[4] for Americans, then the optimum ratio of hedged to unhedged foreign currency claims or liabilities can be derived from a capital asset pricing model. The composition of an optimum portfolio of risky assets less liabilities depends upon $x_1(\mathrm{t})$ the hedged yield, $U(\mathrm{t})$ the unhedged yield, and the variance-covariance matrix.[5]

The *U.S.* demand for a stock of *net* short-term foreign currency claims depends upon the yields (relative to r_H) on hedged and unhedged foreign currency instruments in an optimum portfolio of risky assets, risk aversion, and wealth.[6] At any time, there is

---

[4] Investment in foreign currency assets, or the incurring of liabilities in foreign currency, is riskier than dealing in the domestic currency even when the position is hedged in the forward market. See Lawrence Officer and Thomas Willett, and my 1962 paper. (a) If the investor in foreign currency claims wishes to liquidate his asset at time τ prior to maturity at $t+1$, he must repurchase the forward exchange at price $q_{t+1}(\tau)$ and sell the spot exchange (derived from the sale of the asset) at price $p(\tau)$. His exchange risk is that the forward price may change by a different amount than does the spot price. A similar argument applies to the case where a borrower of foreign currency, who has hedged by purchasing forward foreign exchange, wishes to repay the loan prior to maturity. The exchange risk in hedging does not refer to the price of foreign exchange, as in the unhedged case, but to changes in the forward price relative to the spot price. (b) There is the probability, particularly in times of political or economic instability, that the foreign government may impose restrictions on capital movements. When restrictions are placed upon the sale of foreign exchange for dollars, foreign currency assets are not as liquid as domestic currency assets.

[5] Let $U(\mathrm{t})$ be the expected unhedged return and $x_1(\mathrm{t})$ be the expected hedged return, relative to the safe domestic asset yielding r_H, as defined in Table 1. Then, the expected return per unit of wealth on the portfolio is μ when fraction ξ_u of wealth is invested in the unhedged risky asset and ξ_H in the hedged risky asset $(\sigma_u^2 > \sigma_H^2 > 0)$.

$$\text{(a)} \qquad \xi_u U(\mathrm{t}) + \xi_H x_1(\mathrm{t}) + r_H = \mu$$

The variance of the return on the portfolio is

$$\text{(b)} \qquad \sigma^2 = \xi_u^2 \sigma_u^2 + \xi_H^2 \sigma_H^2 + 2\xi_u \xi_H \sigma_{uH}$$

where σ_u^2 is the variance of p^*, the unhedged return; σ_H^2 is the variance of the hedged return (if the asset is liquidated prior to maturity), and σ_{uH} is the covariance. In an efficient portfolio of risky assets:

$$\text{(c)} \qquad \frac{x_1(\mathrm{t})}{U(\mathrm{t})} = \frac{\xi_H \sigma_H^2 + \xi_u \sigma_{uH}}{\xi_u \sigma_u^2 + \xi_H \sigma_{uH}}$$

If the covariance σ_{uH} is zero, then the ratio ξ_H/ξ_u is described by

(d)

$$\frac{\xi_H(\mathrm{t})}{\xi_u(\mathrm{t})} = \frac{x_1(\mathrm{t})/\sigma_H^2}{U(\mathrm{t})/\sigma_u^2} = \textit{ratio of hedged to unhedged net foreign currency claims}$$

There is good reason to believe that $\sigma_{uH} \approx 0$ (see my 1979 paper, part IV).

Equation (d) implies that the ratio of hedged to unhedged net foreign currency claims is positively related to the expected yield per unit of variance. Using (d), we know the optimum composition of a portfolio of net foreign currency claims; and, given risk aversion and total wealth, we can determine the total demand for net foreign currency claims as described by equation (1).

[6] A firm doing business abroad often derives a marginal convenience yield from holding foreign currency deposits as working balances. These working balances economize on transactions costs. Even if the hedged and unhedged yields were zero, there could be a positive stock of net foreign currency claims that would be demanded. Similarly, there could be a net stock of foreign currency liabilities that firms may want to maintain even when $x_1(\mathrm{t}) = U(\mathrm{t}) = 0$, to economize on transactions costs. The marginal convenience yields and wealth effects are subsumed under variable $a(\mathrm{t}) \gtreqless 0$ in equation (1) below.

a stock $S(t)$ of net claims against foreigners. To avoid revaluing the stock whenever the price of foreign exchange varies, it is convenient to assume that the foreigners neither held dollar assets nor have they incurred dollar liabilities. Then the stock of net short-term claims against foreigners $S(t)$ is just the *net* foreign currency claims.

Equation (1) states that the price of foreign exchange $p(t)$ adjusts immediately to equate the demand for a stock of net short-term foreign currency claims to the stock in existence, $S(t)$. This stock can be positive or negative, depending on whether short-term foreign currency claims are greater than or less than short-term foreign currency liabilities.

$$(1) \qquad A[U(t), x_1(t), a(t)] = S(t), \quad A_i > 0$$

It is convenient for subsequent algebraic analysis to linearize equation (1) as (1′), using the definitions of the uncovered and covered yields on short-term foreign currency claims relative to claims on dollar assets. Risk aversion is reflected by $1/a_1$ and $1/a_2$, the required changes in the uncovered and covered yields, respectively, to induce a unit increase in the stock of *net* short-term foreign currency claims demanded. Parameter a_0 reflects $a(t)$; the effects of the marginal convenience yield, risk, and wealth upon the demand for net foreign currency claims.

$$(1') \quad a_0 + a_1[Ep^* - p(t) - x_3(t)] + a_2[x_2(t) - x_3(t)] = S(t)$$

A change in the stock of net short-term foreign currency claims $\Delta S(t)$, resulting from a nonzero basic balance, produces a change in the spot price $\Delta p(t)$ to induce a short-term capital flow $\Delta A(t)$ which equilibrates the stock demand with the new stock in existence $S(t) + \Delta S(t)$. Equation (1) or (1′) can be viewed as determining the price of foreign exchange which equates the demand $A(t)$ for a stock of net short-term foreign currency claims with the stock in existence $S(t) \gtreqless 0$. Or, the change in the price of foreign exchange induces a short-term capital flow $\Delta A(t)$ which offsets the basic balance $B(t) = \Delta S(t) = \Delta A(t)$.

B. *Forward Market Equilibrium*

The demand for a stock of hedged short-term foreign currency claims depends upon the product of (i) the ratio of hedged to total foreign short-term currency claims in an optimal portfolio of risky assets, and (ii) the total stock of short-term foreign currency claims demanded. The ratio (i) of hedged to total depends positively upon the hedged return $x_1(t)$ and negatively upon the unhedged return $U(t)$. The total stock demand (ii) depends positively upon both the hedged and unhedged returns as described in the previous section. It follows that the stock demand for hedged foreign currency claims is positively related to the net covered yield $x_1(t)$; but, there may be a weak relation between the uncovered yield $U(t)$ and the total stock of hedged foreign currency claims demanded, since a rise in $U(t)$ affects (i) and (ii) in opposite directions. Similarly, the stock of hedged short-term foreign currency liabilities that people wish to incur depends negatively upon the hedged cost $x_1(t)$; and there may be a weak relation between the unhedged cost $U(t)$ and the total stock of hedged foreign currency liabilities for the same reason that applies to assets.

Consequently, the *net* stock of *hedged* short-term foreign currency claims demanded is primarily a function of the net covered yield on foreign currency assets: $x_1(t) = x_2(t) - x_3(t)$, the forward rate less the interest rate differential. This net stock demand for hedged foreign currency claims is precisely the net stock supply of forward foreign exchange by hedgers, denoted by H in equation (2). If $x_1(t)$ is positive, there will be a net stock supply of forward foreign exchange offered by hedgers, $H > 0$. When $x_1(t)$ is negative, then there is a net stock demand for forward foreign exchange by hedgers (for example, by firms which have chosen to borrow foreign currency with the exchange risk covered), H is negative.

$$(2)\quad H(t) = H[x_1(t)] = H[x_2(t) - x_3(t)]$$

= *net stock supply of forward exchange offered by hedgers*

Since foreign investment is riskier than domestic investment, many firms do not have a perfectly elastic supply of a stock of forward foreign exchange at a net covered yield of zero. To induce firms and households to increase the quantity demanded of hedged net foreign currency claims, the net covered yield must rise. Mathematically, this discussion implies that H' in equation (2) is not infinite over a wide range, though it may be infinite at $x_1(t)=0$ over a narrow interval.

On the other side of the forward market are speculators, *defined* as those who do not have offsetting spot and forward positions. The net stock of forward exchange demanded by speculators is positively related to the expected return—the percentage difference between the spot price expected to prevail at the maturity of the contract and the current price of the forward foreign exchange. Let Ep^* denote the logarithm of the expected price of foreign exchange at the maturity of the contract at t+1, and $q_{t+1}(t) = p(t)+x_2(t)$ be the logarithm of the forward price of foreign exchange for delivery at t+1. The forward premium or discount is $x_2(t)$. The net speculative stock demand for forward exchange, denoted by $G \gtreqless 0$, is positively related to the anticipated return $Ep^* - q_{t+1} \equiv Ep^* - p(t) - x_2(t)$.

$$(3)\quad G(t) = G[Ep^* - p(t) - x_2(t)]$$

When G is positive, then speculators are long forward foreign exchange; when G is negative, then speculators are short forward foreign exchange.

Equilibrium occurs in the *forward market* when equation (4) is satisfied—the excess stock of forward exchange offered (demanded) by hedgers is equal to the excess stock of forward exchange demanded (offered) by speculators. In this equation H and G have been piecewise linearized as h and g, respectively, to facilitate subsequent algebraic solution.

$$(4)\quad H(t) = h[x_2(t) - x_3(t)] = G(t) = g[EP^* - p(t) - x_2(t)]$$

For relatively small positions, hedgers may have perfectly elastic excess supply functions at $x_1(t)=0$, i.e., $H'=h$ may be infinite. In that range, the forward rate $x_2(t)$ will equal the interest parity $x_3(t)$. At larger positions, the excess supply (H) function may not be perfectly elastic, $H'=h$ may be finite.

C. *The Rate of Change of Net Short-Term Foreign Currency Claims*

For simplicity, it was assumed that short-term claims against, or liabilities to, foreigners are held in foreign-currency-denominated instruments. Then, in a free exchange market, the change in $S(t)$ (the *net* stock of short-term foreign currency claims) is equal to the basic balance, defined as the current account plus the long-term capital account plus unilateral transfers. Thus $B(t) = DS(t)$, $D \equiv d/dt$.

Write the basic balance as a function of the difference between the logarithm of the current price of foreign exchange $p(t)$ and the logarithm of its longer-run equilibrium value $y(t)$, plus a random disturbance term $u(t)$. Let $E[u(t)]=0$ and $Eu(s)u(t) = \delta_{st}\sigma^2$, where δ is the Kronecker *delta*.

$$(5)\quad DS(t) = B(t) = c[p(t) - y(t)] + u(t)$$

Parameter $y(t)$, which is defined as the equilibrium rate of exchange, can be related to the rate of exchange implied by the monetary approach to the balance of payments.[7] Within that framework, the *BOP* will be in equilibrium when (a) the excess demand for money is zero, and (b) relative purchasing power parity prevails. Condition (a) implies that the stock of real balances M/Z_H is equal to the demand $L(\cdot)$, where M is the stock of money and Z_H is the domestic price index.

---

[7] See Lance Girton and Don Roper, Rudiger Dornbusch, Jacob Frenkel, Michael Connolly and J. D. da Silveria, and Jürg Niehans.

CONDITION (a): $M/Z_H=L(\cdot)$

Relative purchasing power parity Condition (b) is satisfied, if the ratio of the domestic price index Z_H to the foreign price index Z_F is equal to the dollar price of foreign exchange.

CONDITION (b): $P=Z_H/Z_F$

When both conditions are simultaneously satisfied, then the price of foreign exchange P is equal to its equilibrium value Y as described by

CONDITION (c): $Y=M/L(\cdot)/Z_F$

Write Condition (c) in terms of logarithms, denoted by lowercase letters: $y(t)=m(t)-z_F(t)-l(t)$. Basic balance equation (5) can then be written as (5′):

$$\text{(5}'\text{)}\quad \begin{aligned} DS(t)=B(t)&=c[p(t)-y(t)]+u(t)\\ &=c[(p(t)+z_F(t)-z_H(t))\\ &\quad+(z_H(t)+l(t)-m(t))]+u(t)\end{aligned}$$

When the price of foreign exchange $p(t)$ differs from its equilibrium value $y(t)$ in equation (5), then either (i) purchasing power parity is violated, $p+z_F-z_H\neq 0$; or (ii) the stock of money is not equal to the quantity demanded, $z_H+l-m\neq 0$. In case (i), there will be an imbalance in the current account, and in case (ii) there will be an imbalance in the long-term capital account. The net result is that when $p(t)>y(t)$, then there will be on average a positive basic balance $DS(t)>0$, and when $p(t)<y(t)$, there will be on average a negative basic balance.

Coefficient c reflects the speed at which the basic balance responds to deviations of the exchange rate from purchasing power parity and to disequilibrium in the money market. The dynamic process discussed below depends upon the nature of the basic balance function (5). If it contained more lagged terms, the dynamics would be richer but more complex.

II. The Stock Demand Function and Rational Expectations

A. *A General Stock Demand Function Implied by the Model*

When people willingly hold the existing stock of net short-term foreign currency claims (equation (1′) is satisfied) and the forward market is in equilibrium (equation (4) is satisfied), then the resulting spot price is described by equation (6). This is a general stock demand function relating $p(t)$ negatively to $S(t)$. The interest rate differential $x_3(t)$ is exogenous to this model and the price anticipated to prevail in t+1 given the information used by the market in t, denoted by $Ep^*(t+1;t)$, is a parameter in this equation. Stock demand function (6) per se says nothing about market efficiency.

$$\text{(6)}\quad p(t)=-\gamma[S(t)-a_0]+Ep^*(t+1,t)-x_3(t)$$

where $\gamma\equiv(g+h)/[a_1(g+h)+a_2g]$

It is illuminating to rewrite (6) as (6′) where the left-hand side is the expected uncovered yield on foreign currency claims (or the cost of uncovered foreign currency liabilities):

$$\text{(6}'\text{)}\quad \begin{aligned} U(t)&\equiv Ep^*(t+1,t)-p(t)-x_3(t)\\ &=\gamma[S(t)-a_0]\end{aligned}$$

The logic of these equations is quite simple. Suppose that the adjusted stock, defined as $S(t)-a_0$, were positive. Then the price of foreign exchange must induce people to hold this predetermined stock in their portfolios. The expected rate of return on foreign currency claims must exceed the rate of return on domestic currency claims by an amount $U(t)=Ep^*-p(t)-x_3(t)$ which compensates them for the higher risk associated with the claims denominated in foreign currency. The risk premium per unit of foreign exchange, $\gamma[S(t)-a_0]$, rises linearly with the total position and γ reflects risk aversion. In the special case where there is a perfectly elastic supply of arbitrage funds (i.e., h is infinite) then the term γ is $1/a_1$. This term reflects risk aversion since it is the change in the expected unhedged return necessary to induce a unit change in the stock demand (see equation (1′) above).

If $S(t)-a_0$ is negative (for example, there are net short-term foreign currency liabilities), the expected cost must be less than the cost of incurring liabilities denominated in the domestic currency in order to induce people to maintain this stock of net short-term foreign currency liabilities. The reason is that foreign currency liabilities are riskier to *U.S.* residents than corresponding liabilities in domestic currency (just as foreign currency assets are riskier than corresponding domestic currency assets). The relative cost of foreign currency liabilities $U(t)$ should be negative.

It is convenient to define the equilibrium stock of net foreign currency claims denoted by $S_e(t)$ as the stock $S(t)$ associated with no anticipated change in the price, i.e.,

$$(7a)\quad S_e(t)\equiv[S(t)\,|\,Ep^*(t+\Delta t;t)-p(t)=0]$$

which, along stock demand function (6), implies:

$$(7b)\quad \gamma[S_e(t)-a_0]+x_3(t)=0$$

Using the definition of the equilibrium stock $S_e(t)$, stock demand function (6) can be written as

$$(8)\quad p(t)=Ep^*(t+\Delta t;t)-\gamma[S(t)-S_e(t)]\Delta t$$

$$\gamma\equiv(g+h)/[a_1(g+h)+a_2g)]$$

B. *A Rational Expectations Stock Demand Function*

When markets are efficient, then expectations are (rational) informed predictions of future events based upon the true model. The subject of this section is the derivation of a particular stock demand function,[8] when expectations are rational, from the general stock demand function, equation (8). The rational expectations assumption is

$$(9)\quad p^*(t+\Delta t;t)=p(t+\Delta t)+\eta$$

where η has a zero expectation and is not serially correlated.

Let the $V(\cdot)$ function in equation (10) below be the unknown rational expectations stock demand function relating the (logarithm of the) spot price $p(t)$ to the stock $S(t)\gtreqless 0$ of net short-term foreign currency claims.

$$(10)\quad p(t)=V[S(t)]=V_0+V_1S(t)$$

Rational expectations assumption (9) and equation (10) imply that the price expected to prevail at $t+\Delta t$ depends upon the stock expected to be in existence at that date.

$$(11)\quad Ep^*(t+\Delta t;t)=EV[S(t+\Delta t)]$$

Substitute (10) and (11) into stock demand equation (8) and derive:

$$(12)\quad EV[S(t+\Delta t)]-V[S(t)]=\gamma[S(t)-S_e(t)]\Delta t$$

where the length of the period is explicitly introduced. Equation (5) implies that

$$(5'')\quad S(t+\Delta t)\approx S(t)+\{c[p(t)-y(t)]+u(t)\}\Delta t$$

As is shown in Appendix B below, the negatively sloped rational expectations stock demand function (10) is described by equation (13).

$$(13)\quad p(t)=y(t)-\sqrt{\gamma/c}\,[S(t)-S_e(t)]$$

Stock Demand

and is graphed in Figure 1A.

---

[8] The inspiration for this approach comes from Martin Beckmann.

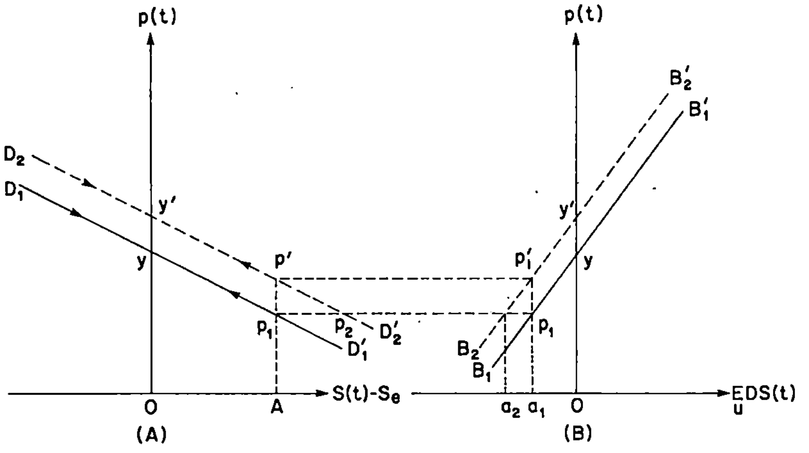


FIGURE 1

III. Exchange Rate Dynamics When Expectations are Rational

A. *Trajectories of the Spot Price and Stock of Net Short-Term Claims on Foreigners*

The dynamics of the spot price of foreign exchange, in a free exchange market with rational expectations, are described by equations (5) and (13) and are graphed in Figures 1A-B and 2.

At any time, the stock of net short-term foreign currency claims $S(t)$ is predetermined. The spot price $p(t)$ adjusts almost immediately to equate the quantity demanded to the stock in existence. The short-term capital flow exactly offsets the basic balance. This price is determined by stock demand equation (13), graphed as DD' in Figure 1A, so that the rationally anticipated percentage change in price is equal to the marginal opportunity cost plus a risk premium. If the stock $S(t)-S_e(t)$ is equal to $0A$ in Figure 1A, when the flow equilibrium price is $y(t)$, then $p(t)$ will be p_1. The slope of the stock demand function (Figure 1A) depends upon risk aversion and the responsiveness of the basic balance to the spot price.

Given the flow equilibrium price $y(t)$ and random shocks $u(t)$, the spot price $p(t)$ determined by the stock demand function D_1D_1' operates upon the basic balance (equation (5) or Figure 1B) to determine the change in net short-term foreign currency claims. To continue with the above example, if the spot price is p_1 then (on average) the basic balance (i.e., the change in the stock of net foreign currency claims) is $DS(t)=0a_1<0$ in Figure 1B. The new stock of net short-term foreign currency claims is $0A+0a_1\cdot\Delta t<0A$. Given the smaller quantity $S(t)$, the new spot price will exceed p_1. The dynamics are indicated by the arrows in Figure 1A. The convergence of $S(t)$ to its equilibrium value S_e is derived by substituting (5″) into (13) to derive equation (14).

$$(14)\quad DS(t)=-\sqrt{\gamma c}\,\left[S(t)-S_e(t)\right]+u(t);$$

$$D\equiv d/dt$$

The left-hand side is the basic balance, which is the excess flow supply of goods and long-term securities in a free exchange market. Equation (14) states that if the stock of net short-term foreign currency claims exceeds the equilibrium value (i.e., there is an "excess" stock $S(t)-S_e(t)>0$), then there will be an excess flow demand for goods and long-term securities (i.e., a negative basic balance). It is for this reason that $S(t)$ converges monotonically to $S_e(t)$ in the absence of shocks.

To determine algebraically the trajectory of the spot price $p(t)$, differentiate stock

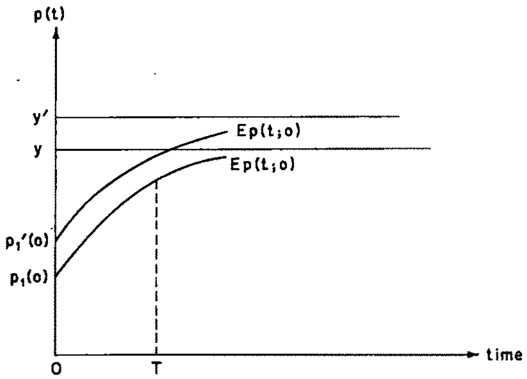


FIGURE 2

demand equation (13) with respect to time and substitute the basic balance equation (5′) for DS(t). Equation (15) is derived. Assume that the equilibrium stock S_e(t), defined in (7b), is constant.

$$(15)\quad Dp(\mathrm{t}) = Dy(\mathrm{t}) - \sqrt{\gamma c}\,[\,p(\mathrm{t}) - y(\mathrm{t})\,] - \sqrt{\gamma/c}\;u(\mathrm{t});\; Eu(\mathrm{t}) = 0$$

$$D \equiv d/d\mathrm{t}$$

This equation describes the percentage change in the spot price (since p(t) is a logarithm) in a rational expectations framework. It is my basic equation.

We can solve for p(t) in closed form and derive equation (16a) or (16b). The term in braces is the rationally expected price at time t from the vantage point of t=0. When flow equilibrium price y(t) is constant, then equation (16a) is described by Figure 2.

$$(16\mathrm{a})\quad p(\mathrm{t}) = \left\{y(\mathrm{t}) + [\,p(0) - y(0)\,]\,e^{-\sqrt{\gamma c}\,\mathrm{t}}\right\} - \sqrt{\gamma/c}\int_0^{\mathrm{t}} e^{-\sqrt{\gamma c}\,(\mathrm{t}-\tau)}u(\tau)\,d\tau$$

$$(16\mathrm{b})\quad p(\mathrm{t}) = Ep(\mathrm{t};0) - \sqrt{\gamma/c}\int_0^{\mathrm{t}} e^{-\sqrt{\gamma c}\,(\mathrm{t}-\tau)}u(\tau)\,d\tau$$

The spot price p(t) converges asymptotically to flow equilibrium price y(t), except insofar as it is subjected to shocks u(t) associated with the basic balance of payments. In Figure 1A, the spot price is determined by the stock demand function DD', equation (13), and the stock of net short-term foreign currency claims. In Figure 1B, the rate of change of the stock is determined by the price generated by Figure 1A, relative to the flow equilibrium price y(t) and shocks u(t). Figure 2 shows the trajectory of the expected price of foreign exchange at time t, from the vantage point of time zero. The expected price Ep(t; 0) converges asymptotically to flow equilibrium price y.

B. *Effects of Disturbances Upon Spot Price Dynamics*

1. *A Parametric Change in the Price which Produces a Zero Basic Balance*

Suppose that flow equilibrium price y(t), rose from y to y' because either the stock of money rose relative to the demand or the foreign price level declined. The basic balance function (equation (5)) shifts upwards in Figure 1B from B_1B_1' to B_2B_2', with a new intercept y' associated with a zero basic balance of payments (flow equilibrium). If the price of foreign exchange remained at p_1, the basic balance deficit would have increased to a_2 from a_1.

The rational expectations stock demand function (equation (13)) shifts up vertically in Figure 1A from D_1D_1' to D_2D_2' with a new intercept of y'. The rise in the rationally expected price of foreign exchange raises the unhedged yield on the stock of net foreign currency claims, and increases the stock of net foreign currency claims demanded. Given the stock of net short-term foreign currency claims *0A* (Figure 1A), there is an excess demand equal to p_1p_2 at price p_1; and the price of foreign exchange rises immediately to p_1' to equate the new stock demand D_2D_2' with the existing stock *0A*.

No change occurs in the basic balance, because the price of foreign exchange immediately adjusts to the change in the flow equilibrium price. The change in the spot price p_1p_1' is equal to the change in the flow equilibrium price[9] yy'. This sudden rise in

---

[9] This is consistent with the results of Girton-Roper and Connolly-da Silveira.

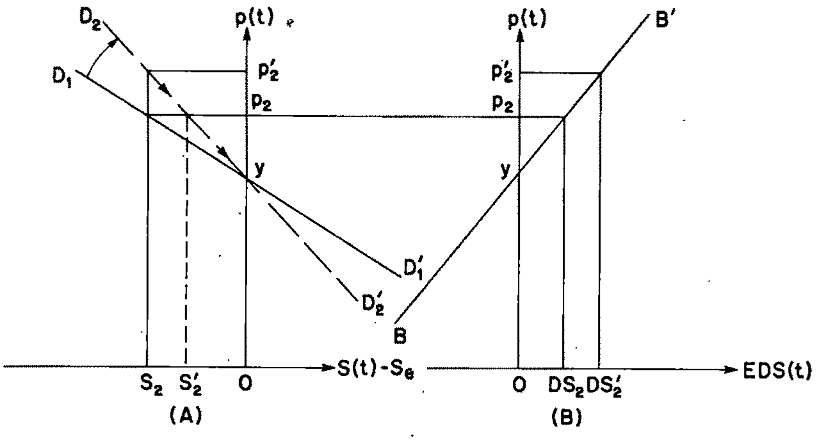


FIGURE 3

price would appear strange to some observers who would note a depreciation of our exchange rate $(p_1p'_1)$ with no significant deterioration in the basic balance.[10] Insofar as old price p_1 was below old flow equilibrium price y due to risk aversion, the new price of foreign exchange p'_1 is below the new flow equilibrium price y'; and there still is a negative balance $0a_1$.

As a result of the basic balance deficits, the stock of net foreign currency claims declines below $0A$. The smaller stock of net foreign currency claims decreases the risk premium $\gamma[S(t)-S_e]$; and the price of foreign exchange converges along the new stock demand function $D_2D'_2$ to y' in the direction of the arrows in Figure 1A. This is described in Figure 2. Moreover, the rise in the price of foreign exchange from p'_1 to y' reduces the basic balance deficit to zero in Figure 1B. Then there is both stock and flow equilibrium.

2. *A Rise in Risk Aversion*

Suppose that there is an increased aversion to risk, such that a larger increase in the expected return is required to induce people to take an uncovered position. This could be produced, for example, by an increase in the variance of disturbance term u or by a greater fear of erratic government intervention in the foreign exchange market. Mathematically, the greater risk aversion is reflected by a rise in γ, the slope of the stock demand function equation (13). Figures 3A-B and 4 describe the effects of a rise in risk aversion upon the spot price dynamics. A rise in γ rotates the stock demand from $D_1D'_1$ to $D_2D'_2$ in Figure 3A. To induce the market to hold the same position, a greater expected return must be offered.

Suppose that there is a net stock of foreign currency liabilities $S_2<0$. Then, a rise in risk aversion would decrease the stock of foreign currency *liabilities* that people want to maintain to S'_2, at price p_2 in Figure 3A. People would try to decrease their outstanding foreign currency liabilities by demanding spot foreign exchange; and thereby repaying their loans. The excess demand for foreign exchange would raise the price to p'_2. A greater expected appreciation of the dollar, or expected depreciation of the foreign exchange, is required to induce people to maintain stock S_2 of net short-term foreign currency liabilities.

As a result of the rise in the price of foreign exchange to p'_2, the basic balance rises from DS_2 to DS'_2. There would be a more rapid convergence of the price $p(t)$ to the flow equilibrium price y.

Figure 4 describes the trajectory of the expected price when the initial stock of net foreign currency liabilities is S_2 (in Figure 3A). A rise in risk aversion from γ to γ' shifts the trajectory from $Ep(t;0;\gamma)$ to

---

[10] This is an implication of rational expectations.

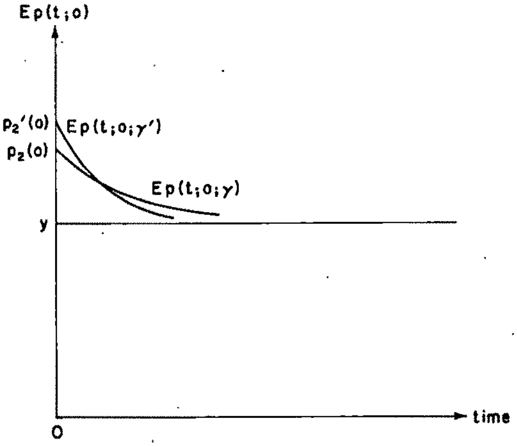


FIGURE 4

$Ep(\mathrm{t}; 0; \gamma')$ in Figure 4. The initial spot price rises from $p_2(0)$ to $p_2'(0)$, the variance of the spot price increases and the expectation of the spot price converges at a faster rate to flow equilibrium y. The speed of convergence is $(\gamma c)^{1/2}$, the product of risk aversion and the sensitivity of the basic balance to the spot price. In Figure 3A, a rise in risk aversion makes the stock demand rotate to D_2D_2', from D_1D_1' to increase its steepness. In Figure 3B, I show the basic balance function relating the expected basic balance to $p(\mathrm{t})-y(\mathrm{t})$. In Figure 4, the trajectory of $p(\mathrm{t})$ to flow equilibrium y is changed when risk aversion rises to γ' from γ.

IV. Efficient Market Tests

A. *Changes in the Spot Price*

The standard "efficient market" tests used in connection with the stock market are not applicable to the foreign exchange (or any other) market where there is a feedback from the price, which equates the stock demand to the stock in existence (equation (13)), to the rate of change in the stock (basic balance equation (5)). In this section Proposition 1 is proved.

PROPOSITION 1: *In a market with rational expectations* (i) *if there exist both risk aversion* ($\gamma > 0$) *and a responsiveness of the basic balance to the spot price* ($c > 0$), *or* (ii) *if the flow equilibrium price* $y(\mathrm{t})$ *is changing, then the percentage change in the spot price adjusted for interest rate differentials will not be a fair game.*[11]

Equation (15) is the rational expectations percentage change in the spot price, assuming that the interest rate differential is constant. The mathematical expectation of $Dp(\mathrm{t})$, with respect to disturbance $u(\mathrm{t})$ is

$$(17)\quad E_u Dp(\mathrm{t}) = Dy(\mathrm{t}) - \sqrt{\gamma c}\,[\,p(\mathrm{t}) - y(\mathrm{t})\,]$$

and Proposition 1 above is based upon this equation.

Suppose that the the stock of money is growing at rate μ relative to the demand for real balances. Then the rate of change of the flow equilibrium price $y(\mathrm{t})$ which produces a zero basic balance of payments is[12]

$$(18)\qquad Dy(\mathrm{t}) = \mu(\mathrm{t})$$

The domestic price level will eventually rise at rate μ, relative to the foreign price level.

Substitute (18) into (17) and derive[13] equation (19).

$$(19)\quad E_u Dp(\mathrm{t}) = \mu - \sqrt{\gamma c}\,[\,p(\mathrm{t}) - y(\mathrm{t})\,]$$

There are two reasons, based upon (19), why the change in the spot price is *not* a fair game when expectations are rational.

First, suppose that the flow equilibrium price were constant at y and interest rate differential x_3 were zero. Figures 2 and 4 describe how the spot price $p(\mathrm{t})$ converges to y at rate $(\gamma c)^{1/2}$; they are based upon $EDp(\mathrm{t}) = -(\gamma c)^{1/2}[\,p(\mathrm{t}) - y(\mathrm{t})]$. As long as $p(\mathrm{t})$ is below (above) y, the spot price is rising (declining). Changes in the spot price even when adjusted for interest rate differentials are serially correlated; therefore,

---

[11] See Appendix A for a definition of a fair game.

[12] Let $M/L(\cdot) = e^{\mu t}$ describe the monetary policy. Then $y(\mathrm{t}) = m(\mathrm{t}) - l(\mathrm{t}) - z_F = \mu \mathrm{t} - z_F$ and $Dy(\mathrm{t}) = \mu$.

[13] This analysis does not depend upon the usefulness of the monetary approach to the balance of payments, but uses it as an example of how the spot price $p(\mathrm{t})$ converges to the flow equilibrium price $y(\mathrm{t})$ where the expected value of the basic balance is zero.

$EDp(t)$ is *not* a fair game. The spot price $p(t)$ is *not* a martingale.

The reason why the spot price is not a martingale is that $Ep^* - p(t)$, the expected change in price, is equal to a risk premium γ times the total adjusted net position $S(t) - S_e$; and there is a flow excess supply $DS(t) = c[p(t) - y(t)] + u(t)$, $c \neq 0$. Insofar as $p(t)$ differs from y, the stock varies. In turn, the total risk premium $\gamma[S(t) - S_e]$ varies. That is why $Ep^* - p(t)$ varies in a *systematic* manner, so that the price is not a martingale. In the stock market literature, the stock is assumed to be perfectly inelastic in supply: $c = 0$ so that there is no feedback from the spot price to the change in the stock. Under those conditions, $EDp(t)$ would be zero. The change in the spot price would be a fair game, because the risk premium is not changing in a systematic way.

Second, suppose that the flow equilibrium price is changing in a systematic way, for example, $Dy(t) = \mu \neq 0$ in equation (19). Then there is a change in the intercept in Figures 1A and 2; and this movement of $y(t)$ is systematic. The spot price moves along the curves in Figure 2 which are shifting in a systematic manner, and are asymptotic to constantly changing values of $y(t)$.

The expected percentage change in the spot price would be given by (19); and that is not a fair game $EDp(t) \neq 0$. Insofar as the interest rate differential $x_3(t)$ reflects the relative rates of monetary expansion, then the percentage change in the spot price adjusted for interest rate differentials is $EDp(t) - \mu = -(\gamma c)^{1/2}[p(t) - y(t)]$, which is not a fair game.

For these two reasons, the change in the spot price (even adjusted for interest rate differentials) will not be a fair game in a foreign exchange market with rational expectations. This proves Proposition 1 above.

There are many types of "nonefficient" markets where expectations are not informed predictions based upon the true model. The question is: Can a nonefficient market be distinguished from an efficient one?

If expectations were formed adaptively, rather than "rationally," then the spot price is described by equations (5), (8), and by (9′) instead of (9).

$$(9') \quad Dp^*(t) = \beta[p(t) - p^*(t)]$$

The dynamic system is

$$(20) \quad Dp(t) = -\gamma c[p(t) - y(t)] - \beta\gamma[S(t) - S_e] - \gamma u(t)$$

$$(5) \quad DS(t) = c[p(t) - y(t)] + u(t)$$

This system is stable, but may display oscillatory behavior if β is large relative to γc. Suppose that $y(t)$ were constant. The equation for the expected change in the spot price:

$$(21) \quad E_u Dp(t) = -\gamma c[p(t) - y(t)] - \beta\gamma[S(t) - S_e]$$

is quite similar to the rational expectations equation (22) based upon (17).

$$(22) \quad E_u Dp(t) = -\sqrt{\gamma c}\,[p(t) - y(t)]$$

Neither equation implies that $EDp(t)$ is zero. Consequently, the standard efficient market test based upon fair game properties of sequences cannot discriminate between rational and adaptive expectations.

This discussion can be summarized as

PROPOSITION 2: *An examination of the time-series per se of spot foreign exchange rates cannot reveal whether or not the foreign exchange market is efficient.*

B. The Forward Price

Many efficient market tests have been concerned with whether the forward price $q_T(t)$ of foreign exchange to be delivered at time T is a martingale, or if the change $q_T(t+1) - q_T(t)$ is a fair game. It is now proved that this property is neither a necessary nor a sufficient condition for a foreign exchange market with rational expectations.

Equations (1′) and (4) can be solved simultaneously for the spot price $p(t)$ and the forward rate $x_2(t)$. The former is described by equation (6) and the latter is described by equation (23). In both equations the anticipated price $Ep^*(t+1;t)$ is a parameter.

(23)
$$x_2(t) \equiv q_{t+1}(t) - p(t) = x_3(t) + \frac{g}{\Delta}[S(t) - a_0]$$
$$\Delta \equiv a_1(g+h) + a_2 g$$

When the hedging function is perfectly elastic ($h \rightarrow \infty$), then g/Δ is zero, and the forward rate $x_2(t)$ is equal to the interest parity $x_3(t)$. Even if the hedging function is not perfectly elastic, a change in the interest parity produces an equal change in the forward rate.[14]

Substitute equation (6) for $p(t)$ into (23) and derive equation (24) which describes the forward price. Let the maturity date be fixed at time $T > t$.

$$(24) \quad q_T(t) = Ep^*(T;t) - \frac{h}{\Delta}[S(t) - a_0]$$

The bias of the forward price from the anticipated price is a function of the total risk premium, which is the product of the adjusted stock of net foreign currency claims $S(t) - a_0 \gtreqless 0$ and risk aversion h/Δ, when the hedging function is perfectly elastic $h/\Delta = 1/a_1$.

When expectations are rational, the anticipated price is the mathematical expectation of the price that will prevail at time T conditional upon the information available at *initial date* t. This rationally anticipated price is described by[15] equation (25) or by the curves in Figure 2 for $y(t)$ equal to either y or y'.

(25)
$$E_u p^*(T;t) = y(T) + [p(t) - y(t)]e^{-\sqrt{\gamma c}(T-t)}$$

Substitute the rationally expected price equation (25) into equation (24) and derive the forward price at time t of a contract maturing at time T.

$$(26) \quad q_T(t) = \left\{ y(T) + [p(t) - y(t)] \cdot e^{-\sqrt{\gamma c}(T-t)} \right\} - \frac{h}{\Delta}[S(t) - a_0]$$

The first term on the right-hand side is the rationally expected price $Ep^*(T;t)$ and the second term is the (positive or negative) risk premium.

In general, the change in the forward price of a contract maturing at time T is equation (27), based upon (24).

(27)
$$Dq_T(t) = DE_u p^*(T;t) - \frac{h}{\Delta}DS(t); \; D \equiv d/dt$$

When expectations are *rational*, the change in the forward price $Dq_T(t)$ is obtained by differentiating (26) with respect to time t, and using (5) and (15) above to derive:[16]

(28)
$$Dq_T(t) = Dy(T) - \sqrt{\gamma/c}\, e^{-\sqrt{\gamma c}(T-t)}u(t) - \frac{h}{\Delta}[c(p(t) - y(t)) + u(t)]$$

---

[14] See John Pippenger and George Pedersson and Edward Tower for evidence concerning the elasticity of supply of arbitrage funds, between the United States and Canada, with respect to the interest rate parity.

[15] The price $p(T)$, given *initial* conditions $p(t)$, is derived from equation (16).

$$p(T) = y(T) + [p(t) - y(t)]e^{-\sqrt{\gamma c}(T-t)} - \sqrt{\gamma/c}\int_t^T e^{-\sqrt{\gamma c}(T-\tau)}u(\tau)\,d\tau$$

Take the expectation $Ep(T;t)$ and derive equation (25).

[16] This derivation is as follows.

(a)
$$DEp^*(T;t) = Dy(T) + [p(t) - y(t)]e^{-\sqrt{\gamma c}(T-t)}\sqrt{\gamma c} + e^{-\sqrt{\gamma c}(T-t)}D[p(t) - y(t)]$$
$$= Dy(T) + e^{-\sqrt{\gamma c}(T-t)} \cdot \left\{ D[p(t) - y(t)] + \sqrt{\gamma c}\,[p(t) - y(t)] \right\}$$

The expected change in the forward price of a contract maturing at time T is equation (29) based upon (28) and $Eu(\mathrm{t})=0$.

$$(29)\quad E_u Dq_{\mathrm{T}}(\mathrm{t})=EDy(\mathrm{T})-\frac{hc}{\Delta}[p(\mathrm{t})-y(\mathrm{t})]$$

The change in the price of a futures contract maturing at time T consists of two parts: (a) the change in the total risk premium resulting from the change in net claims, and (b) the change at time t in the anticipated flow equilibrium price at time T. Consider each component in turn.

(a) The risk premium, or bias of the forward price form the anticipated price, is $(h/\Delta)[S(\mathrm{t})-a_0]$. As the spot price $p(\mathrm{t})$ converges to the flow equilibrium price $y(\mathrm{t})$, the stock of net short-term claims against foreigners changes. Thereby, the risk premium or bias changes in a serially correlated manner: $EDS(\mathrm{t})=c[p(\mathrm{t})-y(\mathrm{t})]$. Given the anticipated flow equilibrium price, the forward price will change along with changes in the risk premium.

(b) The anticipated flow equilibrium price at the maturity of the contract at time T is

$$(30)\qquad y(\mathrm{T})=m(\mathrm{T})-z_F(\mathrm{T})-l(\mathrm{T})$$

as defined earlier. Graphically, let y' in Figure 2 above be $y(\mathrm{T})$. In the model presented here, the uncertainty does not refer to the flow equilibrium price $y(\mathrm{t})$, but to the basic balance associated with any $(p(\mathrm{t})-y(\mathrm{t}))$ in equation (5). Consequently, at any time, $EDy(\mathrm{T};\mathrm{t})=0$; and the change in the forward price $Dq_{\mathrm{T}}(\mathrm{t})$ is not a fair game, *only* because the risk premium changes at rate $(h/\Delta)DS(\mathrm{t})$. If the change in the stock is sufficiently small, due to the narrow interval of time considered, then the forward price $q_{\mathrm{T}}(\mathrm{t})$ should indeed be a martingale.[17]

In reality, the flow equilibrium price $y(\mathrm{T})$, which is an *exogenous variable*, is not known with certainty. For example, the expectation of $y(\mathrm{T};\mathrm{t})$ may converge to a true value (say) given by equation (30) as t converges to T. As time approaches the maturity date T, market participants make even more accurate forecasts of the $m(\mathrm{T})-Z_F(\mathrm{T})-l(\mathrm{T})$, which are exogenous to the model. Suppose that y' in Figure 2 corresponds to $y(\mathrm{T})$ in equation (27); but, at the initial date t, the expectation $Ey(\mathrm{T};\mathrm{t})$ is given by y. As t converges to T, the public understands better what the policy and exogenous variables will be, and $Ey(\mathrm{T};\mathrm{t})$ converges to y'. Thus $EDy(\mathrm{T})$ in equation (27) is not zero, but, as illustrated in Figure 2, it is positive.

Formally, it is suggested that the error $\phi(\mathrm{T}-\mathrm{t})$ in forecasting the policy and exogenous variables in force at time T diminishes as t approaches T.

In my current research on commodity markets, I found that the bias and forecast errors decrease as the time to maturity of a forward contract decreases.

(31a)

$$[Ey(\mathrm{T};\mathrm{t})-y(\mathrm{T})]^2=\phi(\mathrm{T}-\mathrm{t});\ \phi(0),\phi'>0$$

Then, the change in the expected flow equilibrium price is

$$(31\mathrm{b})\quad DEy(\mathrm{T};\mathrm{t})=\frac{-\phi'(\mathrm{T}-\mathrm{t})}{2[Ey(\mathrm{T};\mathrm{t})-y(\mathrm{T})]}\neq 0$$

Using (15)

$$Dp(\mathrm{t})=Dy(\mathrm{t})-\sqrt{\gamma c}\,[p(\mathrm{t})-y(\mathrm{t})]-\sqrt{\gamma/c}\,u(\mathrm{t})$$

$$(\mathrm{b})\quad DEp^*(\mathrm{T};\mathrm{t})=Dy(\mathrm{T})-\sqrt{\gamma/c}\,e^{-\sqrt{\gamma c}(\mathrm{T}-\mathrm{t})}u(\mathrm{t})$$

Substitute (b) into (27) and obtain

(c)

$$Dq_{\mathrm{T}}(\mathrm{t})=Dy(\mathrm{T})-\sqrt{\gamma/c}\,e^{-\sqrt{\gamma c}(\mathrm{T}-\mathrm{t})}u(\mathrm{t})-\frac{h}{\Delta}DS(\mathrm{t})$$

Using (5) in (c), derive

$$(\mathrm{d})\quad Dq_{\mathrm{T}}(\mathrm{t})=Dy(\mathrm{T})-\sqrt{\gamma/c}\,e^{-\sqrt{\gamma c}(\mathrm{T}-\mathrm{t})}u(\mathrm{t})-\frac{h}{\Delta}[c(p(\mathrm{t})-y(\mathrm{t}))+u(\mathrm{t})]$$

which is equation (28).

[17] The seminal work is by Paul Samuelson. We could allow new information about $y(\mathrm{T})$ to be a random variable with a zero expectation, so that there is no trend to $Ey(\mathrm{T};\mathrm{t})$.

Substitute (31) into (29) and derive (32).

$$E_u Dq_T(t) = \frac{-\phi'(T-t)}{2[Ey(T;t)-y(T)]} - \frac{hc}{\Delta}[p(t)-y(t)] \tag{32}$$

The forward price will not be a martingale (and its change will not be a fair game) because both the risk premium, and the error in forecasting the policy and exogenous variables in force at time T, vary in systematic ways.

This section may be summarized by

PROPOSITION 3: *In a model of rational expectations the forward price* $q_T(t)$ *will not be a martingale if*: (i) *the spot price affects the basic balance* ($c>0$) *and there is risk aversion* ($\gamma>0$), *or* (ii) *the market participants' forecast errors of the policy and exogenous variables in force at time* T *decrease as the distance to time* T *decreases.*

An implication of the propositions above is that the standard efficient market tests do not have the power to discriminate between "rational" expectations and some other way of forming anticipations. The contribution of this paper is a description of the trajectories of spot and forward exchange rates when expectations are formed rationally, (i.e., are based upon the structure of the model),[18] and there is a feedback from the stocks to the price to the rate of change of the stock.

APPENDIX A: TESTS OF MARKET EFFICIENCY

Concepts: If a price $Z(t)$ is a *martingale*, then the conditional expectation of the price at the next period is the current price, as described by equation (A1). The expectations operator is denoted by E.

$$E[Z(t+1); Z(t), Z(t-1),\ldots, Z(0)] = Z(t) \tag{A1}$$

This means that all information concerning the past history of prices which affect $Z(t+1)$ is fully reflected in the current price $Z(t)$. The best predictor of $Z(t+1)$ is $Z(t)$. Adding $Z(t-1), Z(t-2)\ldots$or any other series to $Z(t)$ will not improve the prediction of $Z(t+1)$.

Sequence $\{Z(t)\}$ is a martingale if, and only if, the change in the price $X(t)\equiv Z(t)-Z(t-1)$ is a fair game. A sequence $\{X(t)\}$ is a *fair game* if (A2) is satisfied.

$$E[X(t+1); X(t), X(t-1),\ldots, X(1)] = E[X(1)] = 0 \tag{A2}$$

This means that the expected change in the price is zero.

If a sequence is a martingale, then no *filter* (i.e., buy and sell) *rule* will be profitable on average. Equation (A3) will be satisfied when $Z(t)$ is a martingale and $\epsilon(t)$ represents a buy (-1) or sell $(+1)$ decision. Variable $W(t)$ represents accumulated wealth resulting from the $\{\epsilon(t)\}$ strategy, that is, filter rule. Equation (A3) states that every sequence of decision functions (filter rule) $\{\epsilon(t)\}$ converts martingale $\{Z(t)\}$ into martingale $\{W(t)\}$.

$$E[W(t)] = W(t-1) + \epsilon(t)E[Z(t)-Z(t-1)] = W(t-1) \tag{A3}$$

See William Feller (pp. 210–13) for a discussion of martingales.

Empirical Tests: To test the martingale hypothesis the researchers have examined whether disturbance term $u(t+1)$ in the following equations is a serially uncorrelated term with a zero expectation: $P(t)$ is the spot price, $Q_T(t)$ is the forward price at time t for delivery at date T, and $r_H(t)$ and $r_F(t)$ represent the home and foreign interest rates, respectively.

$$P(t+1)-P(t) = u(t+1); \tag{A4}$$

or

$$\frac{P(t+1)-P(t)}{P(t)} = u(t+1)$$

---

[18] See John Muth.

(A5) $Q_T(t+1) - Q_T(t) = u(t+1);$

or $$\frac{Q_T(t+1) - Q_T(t)}{Q_T(t)} = u(t+1)$$

(A6)

$$P(t+1) - P(t)\frac{[1+r_H(t)]}{[1+r_F(t)]} = u(t+1);$$

or $$\frac{P(t+1)}{P(t)} - \frac{[1+r_H(t)]}{[1+r_F(t)]} = u(t+1)$$

(A7) $P(t+1) - Q_{t+1}(t) = u(t+1);$

or $$\frac{P(t+1)}{Q_{t+1}(t)} - 1 = u(t+1)$$

Dooley and Jeffrey Shafer examined change in the dollar spot rates during the period March 13, 1973 to September 5, 1975 for Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Switzerland and the United States. Their conclusion is that:

> Substantial evidence...leads us to reject the martingale model for spot exchange rates [for four out of nine countries, at the 95 percent confidence levels]. The profitability of the filter rules suggest that the deviations from a martingale are important and that exchange markets for many currencies may not have been efficient in the use of price information.... Interest differentials must be ruled out as the principal explanation for the deviations from a martingale.... [p. 27 ff.]

Phillip Cummins et al. examined the Canadian-*U.S.* exchange rate in the 1970–74 period. They concluded that the spot market seems to behave efficiently: it does do a random walk. However, their test indicated that the forward rate on the Canadian dollar does not do a random walk, hence, the forward Canadian-*U.S.* dollar market does not pass this usual weak form test of efficiency.

Richard Levich (Table 5) examined equation (A4) for nine countries during the period 1967–75. For the three-month horizon (where the unit of time is one quarter), the error $u(t+1)$ was *not* significantly different from zero in France and Italy; but the errors seem to be significantly different from zero in Canada, the United Kingdom, Belgium, Germany, the Netherlands, Switzerland, and Japan. Only in two out of the nine countries was the martingale hypothesis consistent with the data.

Another test concerns (A6), whether the percentage change in the spot price is equal to the interest rate differential. The hypothesis is that the error $u(t+1)$ is not significantly different from zero. This hypothesis also fares badly. For a three-month horizon (i.e., the unit of time is a quarter), the mean forecasting error was *not* significantly different from zero only in the United Kingdom. The error is apparently significantly different from zero in the other eight countries examined by Levich (Table 5).

David Kaserman examined the *U.S.*-Canadian dollar during the period July 1955 to March 1961 for the relation between the subsequent spot price $P(t+1)$, where the unit of time is one quarter, and the forward price $Q_{t+1}(t)$ at time t. He studied (A7), and concluded that the "...forward rate...underpredicts the spot rate in periods when the spot rate is rising and overpredicts it when the spot rate is falling" (p. 420). The forecast error $u(t+1)$ in (A7) is positively related to $P(t+1) - P(t)$, the change in the spot price.

Levich (Tables 11 and 10) replicated Kaserman's results for a set of eight countries during 1967–75 and considered one, three, and six-month forward rates. With the possible exception of Belgium for the three-month case, and France for the six-month case, he obtained Kaserman's results. The forward price $Q_{t+1}(t)$ is a biased estimate of the subsequent spot price $P(t+1)$. During 1967–75, $u(t+1)$ in (A7) is positively related to the change in the spot price.

Moreover, even during the German hyperinflation period 1921–23, Frenkel (equation (2′)) found that the forward price was a biased estimate of the subsequent spot price.

Using my notation, he obtained (2′) and the standard errors are in parentheses.

(2′)

$$\log P(t) = \underset{(-0.25)}{-0.45} + \underset{(0.03)}{1.09} \log Q_t(t-1)$$

$$D.W.\ 1.89$$

Coefficient 1.09 is three standard deviations away from its hypothesized value of unity. Consequently:

$$\frac{\Delta \log Q_t(t-1)}{\Delta \log P(t)} = \frac{1}{1.09} = 0.92$$

which is consistent with the Kaserman result replicated by Levich.

Appendix B: Derivation of the Rational Expectations Stock Demand Function

Let (A8) be the unknown stock demand function.

$$\text{(A8)} \qquad p(t) = V[S(t)]$$

Then the rational expectations assumption must be (A9).

$$\text{(A9)} \qquad Ep^*(t+\Delta t) = EV[S(t+\Delta t)]$$

The anticipated price prevailing at time $t+\Delta t$ must be a function of the stock that can be expected to prevail at that date.

It was proved in equation (10) above that when the demand for $S(t)$ is equal to the stock in existence:

$$\text{(10)} \qquad Ep^*(t+\Delta t; t) - p(t) = \gamma[S(t) - S_e(t)]\Delta t$$

Substitute (A8) and (A9) into (10) and derive:

$$\text{(A10)} \qquad EV[S(t+\Delta t)] - V[S(t)] = \gamma[S(t) - S_e(t)]\Delta t$$

Since the right-hand side of (A10) is linear, the unknown $V(\cdot)$ function is linear. The reason for this is that (A10) is an identity; and the object is to find the unknown parameters of the $V(\cdot)$ function: V_0 and V_1.

$$\text{(A11)} \qquad V[S(t)] = V_0 + V_1 S(t)$$

Substitute (A11) into (A12) and derive:

$$\text{(A12)} \qquad E[V_0 + V_1 S(t+\Delta t)] - [V_0 + V_1 S(t)] = \gamma[S(t) - S_e(t)]\Delta t$$

From the flow excess supply equation (5) derive (A13) and (A14).

$$\text{(A13)} \qquad \Delta t \cdot DS(t) = S(t+\Delta t) - S(t)$$

$$\text{(A14)} \qquad \Delta t\{c[p(t) - y(t)] + u(t)]\} = S(t+\Delta t) - S(t)$$

Substitute (A14) into (A12) and obtain

(A15)

$$E_u V_1[c(p(t) - y(t)) + u(t)]\Delta t = \gamma[S(t) - S_e(t)]\Delta t$$

The expectation of $u(t)$ is zero, and $p(t)$ is given by (A8) and (A11). Hence,

(A16)

$$V_1 c[V_0 + V_1 S(t) - y(t)] = \gamma[S(t) - S_e(t)]$$

This can be written as

(A17)

$$cV_1^2 S(t) + cV_1(V_0 - y(t)) = \gamma S(t) - \gamma S_e(t)$$

This is an identity in $S(t)$. Equate coefficients and obtain

$$\text{(A18)} \qquad cV_1^2 \equiv \gamma$$

$$\text{(A19)} \qquad cV_1(V_0 - y(t)) \equiv -\gamma S_e(t)$$

If the stock demand function is negatively sloped, or if the dynamic system is stable, we must take the negative root of V_1:

$$\text{(A20)} \qquad V_1 = -\sqrt{\gamma / c}$$

Then

$$V_0 = y(t) + \sqrt{\gamma/c}\, S_e(t) \tag{A21}$$

We have, therefore, derived

$$\begin{aligned} p(t) &= V_0 + V_1 S(t) \\ &= y(t) - \sqrt{\gamma/c}\,[S(t) - S_e(t)] \end{aligned} \tag{A22}$$

as the stock demand function.

REFERENCES

M. J. Beckmann, "On the Determination of Prices on Futures Markets," in Michael J. Brennan, ed., *Patterns of Market Behavior: Essays in Honor of Philip Taft,* Providence 1965.

M. Connolly and J. D. de Silveira, "Exchange Market Pressure in Postwar Brazil," *Amer. Econ. Rev.,* June 1979, *69,* 448–54.

Phillip Cummins et al., *Aspects of Efficiency in the U.S./Canadian Foreign Exchange Market,* U.S. Treasury Conference, Washington 1976.

Michael Dooley and P. Isard, "A Portfolio—Balance Rational—Expectations Model of the Dollar-Mark Rate: May 1973–June 1977," *Int. Finance Disc. Papers,* No. 123, Sept. 1978.

______ **and Jeffery Shafer,** *Analysis of Short-run Exchange Rate Behavior,* U.S. Treasury Conference, Washington 1976.

R. Dornbusch, "The Theory of Flexible Exchange Rate Regimes and Macroeconomic Policy," in Jacob A. Frenkel and Harry G. Johnson, eds., *The Economics of Exchange Rates,* Reading 1978, ch. 2.

William Feller, *An Introduction to Probability Theory and its Applications II,* New York 1966.

Jacob Frenkel, "The Forward Exchange Rate, Expectations, and the Demand for Money: The German Hyperinflation," *Amer. Econ. Rev.,* Sept. 1977, *67,* 653–70.

______ **and Harry G. Johnson,** *The Economics of Exchange Rates,* Reading 1978.

L. Girton and D. Roper, "A Monetary Model of Exchange Market Pressure Applied to the Postwar Canadian Experience," *Amer. Econ. Rev.,* Sept. 1977, *67,* 537–47.

D. L. Kaserman, "The Forward Rate: Its Determination and Behavior as a Predictor of the Spot Rate," Amer. Statist. Assn., Bus. Econ. Statist. Sec. 1973.

S. W. Kohlhagen, "The Behavior of Foreign Exchange Markets–A Critical Survey of the Empirical Literature," *Monog. Series Finance Econ.,* Grad. Sch. Bus. Admin., New York Univ. 1978.

R. H. Levich, "Tests of Forecasting Models and Market Efficiency in the International Money Market," in Jacob A. Frenkel and Harry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies,* Reading 1978.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica,* July 1961, *29,* 315–35.

J. Niehans, "Exchange Rate Dynamics with Stock/Flow Interaction," *J. Polit. Econ.,* Dec. 1977, *85,* 1245–57.

L. H. Officer and T. D. Willett, "The Covered Arbitrage Schedule," *J. Money, Credit, Banking,* May 1970, *2,* 247–57.

G. Pedersson and E. Tower, "On the Long and Short-Run Relationship Between the Forward Rate and the Interest Parity," *J. Macroeconomics,* Winter 1979, *1,* 65–77.

J. Pippenger, "Interest Arbitrage Between Canada and the United States: A New Perspecive," *Can. J. Econ.,* May 1978, *11,* 183–93.

P. A. Samuelson, "Proof that Properly Anticipated Prices Fluctuate Randomly," in Robert C. Merton, ed., *The Collected Scientific Papers of Paul A. Samuelson,* Vol. II, Cambridge, Mass.; London 1972.

Jerome L. Stein, *The Nature and Efficiency of the Foreign Exchange Market,* in *Essays in International Finance,* Princeton Univ., No. 40, Oct. 1962.

______, "Spot, Forward and Future," in Haim Levy, ed., *Research in Finance,* Vol. 1, Greenwich 1979.

Experiments with a Decentralized Mechanism for Public Good Decisions

By VERNON L. SMITH*

Public goods theory, as articulated by Paul Samuelson (1954), Mancur Olson, and many others, has maintained that the distinguishing characteristic of nonexcludability, inherent in public goods, implies the impossibility of optimal public good provision by decentralized (voluntary) mechanisms. Since no one can be excluded from the benefits of the public good, each has an incentive to "free ride"; that is, contribute less than his marginal valuation to the cost of the public good. Consequently, public goods theory has suggested two hypotheses.

H_F: *The free rider hypothesis.*

H_I: *The impossibility of decentralized public good provision.*

Traditional public goods theory has not viewed H_F and H_I as independent hypotheses. The assumption of free rider behavior has led to the conclusion that all decentralized mechanisms would lead to the nonprovision of public goods (for example, Earl Brubaker refers to this as a consequence of the strong free rider hypothesis), or at least to suboptimal quantities of public goods.

Precursory and recent contributions by William Vickrey, Edward Clarke (1968, 1971), Theodore Groves (1969, 1973), and Groves and John Ledyard, while accepting the logic of free riding behavior, have sought solutions to the problem in the form of cost-sharing mechanisms that would give each individual an incentive to contribute to the cost of a public good an amount equal to his marginal private valuation of the public good. Consequently, such mechanisms are incentive compatible, that is, individual's acting out of self-interest support outcomes (Pareto optimality) judged to be desirable by the collective. The idea is to find procedures that harness the self-interest to produce optimal *common* outcomes, just as the traditional theory of private goods exchange has argued that action in the self-interest is like an invisible hand guiding collectives to specialize, then cooperate through market exchange, to provide optimal *private* outcomes. The Clarke-Groves-Ledyard (hereafter, C-G-L) research denies H_I, and hypothesizes:

H_p: *The possibility of designing decentralized adjustment mechanisms for optimal public good decision.*

Traditional public goods theory, while asserting H_F and H_I, has not questioned similarly the invisible hand hypothesis in the provision of private goods. In fact the argument for H_I has leaned on a sharp contrast between private goods, where there is no presumed problem of decentralized decision, and public goods, where the problem is assumed to be inevitable. As noted by Samuelson, there is a fundamental technical difference between private and public goods: by departing from the rule of preference revelation, "... any one person can hope to snatch some selfish benefit in a way not possible under the self-policing competitive pricing of private goods" (1954, p. 389).

It might be supposed from the above discussion that the "new" conventional wisdom in public goods is acceptance of the proposition that although there are very important technical differences between public and private goods, it is nevertheless possible, by applying C-G-L types of ingenuity, to design incentive-compatible decentralized pricing systems for public goods. But such a simplified reduction of these important new contributions is not warranted. Leonid Hurwicz, for private goods, and Ledyard and John Roberts, for public goods, have proved that it is impossible to find a mechanism that provides individually rational

---

*University of Arizona. Research support by the National Science Foundation is gratefully acknowledged. I am indebted to Michael Vannoni for programming the PLATO version of the Auction Mechanism.

Pareto optima and which is also individually incentive compatible. These results assume a broader range of (costless) strategic behavior than the Nash competitive behavior postulated by C-G-L. In summary the current state of theory on these issues appears to be as follows: *Depending upon what one assumes about the strategic behavioral modes of economic agents, the attainment of optimal resource allocation by decentralized pricing systems may be either possible or impossible, but in either case this result does not depend on whether goods are private or public.*

There are many lessons in the literature that have brought us full circle to this state, but none more important than the necessity of learning much more about the behavior of people in institutions that attempt to operationalize the significant concept of incentive compatibility. But abstract theory has done more than simply return to where it started, dramatizing its own dead end in the absence of a better understanding of behavior. Together, the above theoretical contributions have brought a glimpse of possibility, of operational forms as in Vickrey for private goods, and Groves and Ledyard for public goods, and a better understanding of the scientific questions that require investigation.

I. Empirical Tests of Public Goods Theory

There is much casual evidence to suggest that mechanisms exist for the voluntary provision of public goods. Thousands of churches, music halls, libraries, scientific laboratories, art museums, theatres, and other such facilities have been financed by voluntary contributions to fund drives conducted by private societies. If such observations are dismissed as evidence contrary to H_I on the grounds that these are instances of atypical altruism, then there is a serious question as to whether public goods theory consists of any refutable hypotheses. If observations contrary to a theory are to be ruled out on the basis of special explanations not part of the original theory, then we are indeed dealing only with Galbraithian systems of belief, not with scientific propositions.

A more substantive objection to these observations of successful private society fund drives might be that they represent "small" groups, whereas the public good problem is a "large" group problem. But regardless of the size of any finite group, one can always imagine a larger collective, and conclude that the hypothesis has not been disproven. Even in a collective of size two there is incentive to free ride, and underprovide the joint outcome. Consequently, if size is claimed to be a critical parameter in public goods theory, the nature of its effect must be well specified before examining the observations. But suppose we grant that action by private societies is not compelling evidence rejecting H_I; that although the distinction between small and large is not precise, private societies are "clearly" too small to represent a valid test of H_I. Then we must look for examples of voluntary action by larger collectives. Such an example is voting in national elections, which, unlike Australia, is not yet required by law in the United States. The problem of voting is often described as a public goods problem: since no individual voter can have any perceivable effect on the common outcome, and since the act of voting involves some cost (going to the polls, reading and marking the ballot, etc.), it follows, allegedly, that one should observe little or no voting activity. In fact, "remarkably large" numbers of people vote. Is this judged to be evidence contrary to H_I by political economists? No, it is a "paradox," and there is a great literature on the paradox of voting with many distinguished contributions (see Gordon Tullock; George Stigler; J. Ferejohn and M. Fiorina). Once again, a cherished belief remains unshaken, and the observation is explained anecdotally.

I am not arguing that H_I is disproved by the fact that people vote, or the observation that private societies are sometimes successful in building-fund drives; I am arguing for clearer standards of evidence. These field observations are clearly inconsistent with the strong free rider hypothesis; large numbers of people vote, and public goods are being provided by private societies. What we do not know and cannot know from field observation is the optimal quantity of the given public good that was provided. Some fund drives succeed, and the facility is built; others fail. In terms of optimality, should some of the latter have succeeded, or should some of the former have provided larger or better facilities? We cannot know what was optimal without independent information on individual valuations.

This problem appears to be inherent in field studies, but not in controlled laboratory experiments which can provide a unique source of empirical data for testing hypotheses about resource allocation mechanisms. It appears that in field studies the best we can do is to compare mechanism *A* with mechanism *B*. Suppose theory suggests that *A* is incentive compatible, while *B* is not. Then we can test the hypothesis that observations on mechanism *A* come from distributions over a larger quantity of the public good than do observations on mechanism *B*. But we cannot test the hypothesis that observations from *A* come from a distribution whose central tendency is the optimal quantity of the public good. This last hypothesis (as well as the former) can be tested using appropriate laboratory experiments because we can, to a close approximation, induce known preferences or values on decision outcomes (see my 1979 and forthcoming papers), and thereby compute the theoretically optimal quantity of the public good.

A. *Laboratory Experiments*

In my previous experimental research using a partial equilibrium design (no income effects), I found that collectives composed of four, five, or eight members are able, in most replications, to reach agreement to produce optimal or near optimal amounts of a public good under either a version of the Groves-Ledyard mechanism or the auction mechanism (see my 1979 paper). However, under the Lindahl mechanism which (unlike the Groves-Ledyard and auction mechanism) is *not* incentive compatible (i.e., Pareto optimal allocations are not among the set of points that are Nash equilibria), I found that experimental collectives did much worse; specifically, in three replications none of the public good was produced, while in one, a positive quantity equal to one-half the optimum was provided. Hence, the Lindahl mechanism experimental results provided evidence in support of H_F, while the results from the two incentive-compatible mechanisms support H_P by demonstrating that operational procedures exist for the solution of the free rider problem.

The auction mechanism, which will be reviewed in greater detail in Section II, is characterized by collective excludability, unanimity, and budget balance. It exploits the fact that prior to the actual provision of a pure public good, a collective, and therefore each of its members, can be excluded from the good by not providing it. Therefore, the stopping rule requires that (a) the sum of individual bids (contributions) cover the cost of the proposed public good, and, if this occurs, (b) the collective *then* agree unanimously to accept this result. I once thought this was a new mechanism, but actually it is just an extension, generalization, and formalization of the age-old "fund drive" procedure used by many private societies and eleemosynary institutions.

It is my conjecture that it is this "pre-contract, group excludability" (see Brubaker) feature that derives these earlier experimental results (see my 1979 paper) and the results to be reported below. But this is not a certainty, as is suggested, perhaps, by the results of J. Sweeney and G. Marwell and R. Ames.

Sweeny reports the results of experiments in which an individual subject pedaled a bicycle connected to an electric generator. Each individual thought she was in a group of size six with the group goal (public good) to maintain the bright illumination of a "group light" for at least ten minutes and thereby receive partial course credit for her participation. In fact, the experimenter regulated the group light with a rheostat, as well as the subject's six individual lights which enabled her to perceive the contributions of other supposed members of her group. Average pedalling speed (revolutions per minute) was used to measure the individual's contribution to the public good. The results showed that subjects in a large group (where perceived individual effectiveness is nil) withhold a statistically significant part of their contributions in comparison to a small group (perceived effectiveness is large). This supports H_F in the context of a situation in which (it may be presumed) subjects believed they would be excluded from course credit if the group light was not kept sufficiently bright. However, two points must be emphasized: 1) of great significance is the fact that contributions were quite large under all experimental conditions. Mean contributions differed from zero far more significantly than from each other, even where the group light was bright and independent of subject speed and other subjects were contributing very little. Thus mean contributions (across six different treatment conditions) varied from speeds of 170.45 to 203.61, while standard errors varied from 26.12 to 58.89. 2) The conditions of excludability were imprecise, that is, there was no exact measurement which indicated to the subject when the group light was adequately bright to meet the group goal. These considerations suggest that H_F is supported only in a *very weak* form, and perhaps only because there was vagueness as to the conditions under which group excludability would occur. Indeed, given the high mean contributions under all treatment conditions, it seems that the results (a) are consistent with those reported in my 1979 paper for an entirely different experimental paradigm, and (b) do not support the generally held belief that free riding is a serious problem in the provision of public goods.

A more recent experimental study by Marwell and Ames also concludes that only a weak version of H_F is supported. The Marwell-Ames results are of particular interest because traditional public goods theory applied to their experimental paradigm provides a clear incentive to free ride. Each subject allocated an endowment of "tokens" (225 in the equal distribution case) between a "private exchange," yielding a well-defined constant private marginal rate of return in money, and a "group exchange" yielding a well-defined marginal rate of return in money that varied with the group's total contribution. The group exchange required an average contribution of 100 tokens per subject to yield a return better than that in the private exchange, and an average of 200 per subject to yield an optimal allocation. It would appear that the incentive to contribute nothing to the public good is very high. Using collectives of size four and eighty, the mean contribution was 127.6, well below the optimum, but remarkably large by conventional reasoning. Since subjects in the Marwell-Ames experiment made only one investment, it is possible to argue that learning effects in a repeat-play version would produce greater (or why not less?) support for H_F. But the Marwell-Ames results do not contradict those in other experiments (see my 1979 paper) that used repeat play. Hence, it seems that these empirical results are robust under replication and call into question the traditional model of public good decision behavior.

B. *Field Experiments*

Beginning in 1974, approximately 150 station managers in the American network of noncommercial television stations have used a computer-based decentralized mechanism called the Station Program Cooperative (*SPC*) to produce and allocate a substantial portion (nearly half in the 1974–75 season) of the programs broadcast by these stations. The *SPC* procedure uses a series of up to twelve trials, and has selected from 25 to 38 programs from menus of over 200 in each of its operating seasons.

Unlike the laboratory experiments reported above, and the field experiment reported below, the *SPC* (inefficiently, in theory) selectively excludes members who do not contribute their share of cost to a given program. On each trial, each station agent votes "yes" or "no" on each program, based on the cost share allocated to him on the previous trial. For the first trial an initialization procedure is followed. At present, the cost borne by station i for program j is simply

$$C_j B_i / \sum_{k \in S_j} B_k$$

where C_j is the production cost of program j, B_i is the budget (measured by the Community Service Grant awarded the station) of station i, and S_j is the set of stations voting for program j on the previous trial. With a little (but not much) discretionary intervention by the center over the twelve-trial sequence, the process converges in ten to twelve trials.

It is not illegitimate to question the correctness, that is, efficiency, of this process. But it is much more interesting to ask if it works, why it works, and whether it can be improved upon using well-defined criteria. It works in the sense that the participants seem heartily to approve of the process, but think it can be improved. It seems to work in the sense of rapid convergence because of strong individual exclusionary features, but in view of the good results from other experiments (less rich in number of commodities) that do not depend on individual exclusion, it is not certain that this is a necessary feature. Ferejohn, Robert Forsythe, and Roger Noll (pp. 43–57) have investigated some of these issues in a series of laboratory experiments comparing the *SPC* procedure with a "*B* procedure" based on a mechanism proposed by Jerry Green and Jean-Jacques Laffont. The termination condition for both procedures was that either 1) the same selections were made by all subjects on two consecutive trials (after the third), or 2) the trial exceed a "secret" number (always ten). The *SPC* terminated under 1) in every case (five experiments), while the *B* procedure terminated under 2) in every case (eight experiments). In terms of efficiency, neither procedure was clearly superior to the other. Hence, for multiple discrete public goods where individual exclusion is possible, the *SPC* stands as the operational mechanism that would-be designers must beat: It makes allocations no worse, and terminates earlier, than the only other mechanism with which it has been compared.

Peter Bohm has reported the results of a significant Swedish television field experiment which provides no support for H_F even in its weak form. The results are significant because they were particularly unexpected in November 1969 when the experiment was conducted. This was well before any of the ideas in the recent literature on incentive compatibility had become known. But Bohm's data gave him no recourse but to conclude "that the well-known risk for misrepresentation of preferences in this context may have been exaggerated" (p. 111). In retrospect the results are not unexpected because the conditions of the Bohm experiment effectively introduce the collective excludability characteristic of the auction mechanism, and, as I have argued elsewhere (1979), the results were actually to be expected because what was wrong was the theory that predicted free riding in the first place.

In summary, the empirical evidence from laboratory experiments, controlled field experiments, and actual field experience (*SPC*) overwhelmingly supports H_P. Not only is it true that incentive-compatible public good mechanisms are possible, and that there exist operational forms for such mechanisms; it is also true that such mechanisms are fairly easy to invent, they are not complicated and difficult to understand, and they are easy to computerize for large-scale decisions. Furthermore, there is some evidence (see Marwell and Ames) to suggest that the problem of free riding is not severe —in fact, pretty weak—even where there are strong and fairly obvious incentives to withhold contributions to a public good. In those mechanisms that are incentive compatible, perhaps this undercurrent of non-free-rider behavior by some agents helps to offset the effect of strategic manipulation (always potentially present) by other agents.

II. The Auction Mechanism for Public Goods

Consider an economy consisting of one private and one public good. Let consumer agent $i(i=1,2,\ldots,I)$ have an endowment ω_i of private good, and a utility function $u^i(y_i, X)$, increasing and concave in (y_i, X) yielding utility u^i if i retains y_i units of the private good for himself and the collective of I agents chooses to produce X units of the public good. A unit of the public good can be produced from q units (a constant) of the private good.

Let each agent choose a 2-tuple (B_i, X_i) where $B_i=\omega_i-y_i$ is the bid in private good units that i contributes to the production of the public good, and X_i is the quantity of the public good proposed by i. Define:

The collective's proposed quantity of the public good is the mean, $\overline{X}=\Sigma_{\forall k}X_k/I$, of the individual proposals.

The mean of all proposals excluding i is $\overline{X}_i=\Sigma_{\forall j\neq i}X_j/(I-1)$.

The sum of all bids excluding i is $\hat{B}_i=\Sigma_{\forall j\neq i}B_j$.

The net (residual) price of the public good to i is $q-\hat{B}_i/\overline{X}_i$, and his share of the cost of the collective's proposal is $(q-\hat{B}_i/\overline{X}_i)\overline{X}$.

Collective agreement (equilibrium) requires each i to accept his share of cost by bidding that amount and to accept the collective's proposed quantity of the public good as his personal proposal, that is, agreement occurs when there is unanimity in the sense that $B_i=(q-\hat{B}_i/\overline{X}_i)\overline{X}$ and $X_i=\overline{X}, \forall i$. A procedure for operationalizing these conditions is contained in the process described in the next section.

Assume $u^i(\omega_i,0)<u^i(y_i, X)\forall y_i<\omega_i, X>0$. Then i receives utility

$$W_i=\begin{cases} u^i\left[\omega_i-\left(q-\hat{B}_i/\overline{X}_i\right)\overline{X}, \overline{X}\right], & \\ \quad \text{if } B_i=\left(q-\hat{B}_i/\overline{X}_i\right)\overline{X},\ X_i=\overline{X},\ \forall i & \\ u^i[\omega_i,0], \quad \text{if } B_i\neq\left(q-\hat{B}_i/\overline{X}_i\right)\overline{X} & \\ \qquad\qquad \text{or } \overline{X}_i\neq\overline{X} \text{ for any } i & \end{cases} \tag{1}$$

It will be shown that if each i chooses (B_i, X_i) to maximize W_i, the resulting conditions correspond to a Lindahl equilibrium which is also a noncooperative or Nash equilibrium. However, (1) defines many Nash equilibria which are not Lindahl equilibria.

Assuming differentiability, a maximum of W_i requires "agreement" as defined above, and $-(q-\hat{B}_i/\overline{X}_i)(1/I)u_1^i+(1/I)u_2^i=0$, or

$$B_i=\left(q-\hat{B}_i/\overline{X}_i\right)\overline{X},\ \forall i \tag{2}$$

$$X_i=\overline{X},\ \forall i \tag{3}$$

$$\frac{u_2^i}{u_1^i}=q-\hat{B}_i/\overline{X}_i,\ \forall i \tag{4}$$

From (2), since (3) implies $\overline{X}_i=\overline{X}$,

$$B_i+\hat{B}_i=q\overline{X},\ \forall i \tag{5}$$

and, from (4), (2), and (5)

$$\sum_{\forall k}\frac{u_2^k}{u_1^k}=\frac{\sum_{\forall k}B_k}{\overline{X}}=\frac{B_i+\hat{B}_i}{\overline{X}} \tag{6}$$

$$=\frac{\sum_{\forall k}(w_k-y_k)}{\overline{X}}=q$$

which defines an interior $(0<y_i^0<\omega_i)$ Lindahl equilibrium, $(y_1^0,\ldots, y_I^0, X^0)$. This is also a Nash equilibrium since if all $j\neq i$ bid $B_j^0=\omega_j-y_j^0$, and propose $X_j=X^0$, agent i's best choice is $B_i^0=\omega_i-y_i^0$ and $X_i=X^0$; otherwise he prevents agreement and receives utility $u^i(\omega_i,0)<u^i(y_i^0, X^0)$.

That there exist Nash equilibria that do not satisfy (4) and (6) is seen by considering any $(y_1^*,\ldots, y_I^*; X^*)$ such that $u^i(y_i^*, X^*)>u^i(\omega_i,0)\forall i$ and

$$\sum_{\forall i}\omega_i=\sum_{\forall i}y_i^*+qX^*$$

Then if $\forall j\neq i$, $B_j=\omega_j-y_j^*$, $X_j=X^*$, agent i's best choice is to agree to the arrangement by setting

$$B_i=\omega_i-y_i^*=-\sum_{j\neq i}\omega_j+\sum_{j\neq i}y_j^*+qX^*$$

and $X_i=X^*$.

It is well known that a Lindahl equilibrium is among the set of Pareto optimal arrangements, where the latter must satisfy (6). The Lindahl equilibrium defined by (4) and (6) and the Pareto optimal set defined by (6) can be contrasted with the free rider solution. To obtain this solution, assume that each agent i chooses y_i on the assumption that X is determined by whatever budget $\Sigma_{\forall i}(\omega_i - y_i)$ results from these choices. Thus if each i chooses y_i, $0 \leqslant y_i \leqslant \omega_i$, then $X = \Sigma_{\forall i}(\omega_i - y_i)/q$. Each i is then assumed to $\max_{y_i} u^i(y_i, \Sigma_{\forall i}(\omega_i - y_i)/q)$. Hence, at a maximum, $(y_1', y_2', \ldots, y_1'; X')$, we have

$$\frac{u_2^i}{u_1^i} \leqslant \frac{\sum_{\forall i}(\omega_i - y_i')}{X'} = q \tag{7}$$

where $y_i' = \omega_i$ if $<$ holds for any i. In (7) we allow for boundary solutions $y_i' = \omega_i$ since they occur in the experimental design reported in Section III below.

The auction mechanism experiments reported in my 1979 paper correspond to the model stated above if we set $u_1^i \equiv 1$, $\forall i$. The utility function is then $u^i(y_i, X) \equiv y_i + u_i(X)$ and $u_2^i = u_i'(X)$. Consequently, the set of Pareto optimal allocations defined by (6) is simply the quantity of the public good X^0 satisfying $\Sigma_{\forall i} u_i'(X^0) = q$, which is also the Lindahl equilibrium quantity if this equation has a unique solution. The Lindahl prices $b_i = u_i'(X^0)$ are given by (4) with b_i representing agent i's bid in numeraire units per unit of the public good. The budget constraint is necessarily satisfied since $X^0 \Sigma_{\forall i} b_i = qX^0$ identically.

III. PLATO Public Good Experiments

The experiments to be reported here used a computerized iterative decision process to operationalize the mechanism of the preceding section. The computations in this process are too complex to have been executed effectively without computer aid. The PLATO computer system with visual display consoles was programmed to execute the procedures that guide subject groups through the decision process. Such a system has enormous advantages over manually executed experiments in standardizing procedures, eliminating possible experimenter or "auctioneer" effects, recording subject choices, performing routine calculations, and displaying the appropriate private or public information to the subjects.

Prior to the commencement of an experiment it must be initialized by the experimenter. For initialization PLATO requires numerical values for $(I, T, q, \omega_i, a_i, b_i, \alpha_i, \beta_i)$, where I is the number of subject agents who will participate in the experiment, T is the maximum number of decision trials, q is the unit cost of the public good, ω_i is the private good endowment of subject i, and the payoff function in dollars to i is $V^i(y_i, X) \equiv a_i y_i^{\alpha_i} X^{\beta_i} - b_i$ for all (y_i, X) such that $V^i \geqslant 0$.

The language of the process consists of private messages $(B_i(t), X_i(t))$ sent by agent i on trial t to PLATO, and private messages $(q - \hat{B}_i/\bar{X}_i)\bar{X}$ and public messages $(\Sigma_{\forall i} B_i, \bar{X})$ sent by PLATO to each i on each trial. The process uses the following iterative procedure:

(i) *Starting rule*: On trial t each agent, using his or her typewriter keyboard, independently and privately types a 2-tuple $(B_i(t), X_i(t))$ where $B_i(t) \in [0, \omega_i]$ and $X_i(t) \in [0, 12]$. Agent i, at his discretion, then presses a key that actually transmits his selection to the "center," that is, a memory space in PLATO.

(ii) *Transition rules*:

(a) After the last subject-agent has entered his choices, PLATO computes $(\Sigma_{\forall i} B_i, \bar{X})$. In sixteen of the experiments below, this message was transmitted to each agent, while in thirteen experiments, $\bar{X}$ was first rounded to the nearest integer before the message was sent.

(b) If $\Sigma_{\forall i} B_i \geqslant q\bar{X}$, the collective has reached tentative agreement. If $\Sigma_{\forall i} B_i = q\bar{X}$, each agent's share of cost is set equal to his bid $q\bar{X} - \Sigma_{\forall j \neq i} B_j = B_i$. This fact is reported to each i, and PLATO goes into a voting mode in which each agent is asked to type "yes" or "no" indicating whether he wishes to accept or not accept as final the arrangement resulting from that trial. If $\Sigma_{\forall i} B_i > q\bar{X}$, PLATO modifies the collective's decision $(\Sigma_{\forall i} B_i, \bar{X})$ so that the center retains none of the bid surplus. Two procedures were used in this modification:

1. Rebate rule 1: PLATO first determines if the overbid is sufficient to allow an integer increase in the size of the public

good, that is, the largest $n=0,1,2,\ldots$ such that $\Sigma_{\forall i} B_i \geqslant q(\bar{X}+n)$ was chosen. Any remaining bid surplus is then rebated to each agent in proportion to his bid. Thus, if agent k's bid was B_k the adjusted bid becomes

$$B_k' = B_k \left[1 - \frac{\sum\limits_{\forall i} B_i - q(\bar{X}+n)}{\sum\limits_{\forall i} B_i} \right] \tag{8}$$

$$= \frac{B_k q(\bar{X}+n)}{\sum\limits_{\forall i} B_i}$$

where $\bar{X}+n$ is the adjusted proposed size of the public good.

2. Rebate rule 2: PLATO adjusted the outcome $(\Sigma_{\forall i} B_i, \bar{X})$ to give $(\Sigma_{\forall i} B_i', \bar{X}')$ where the vector from $(\Sigma_{\forall i}(\omega_i - B_i), \bar{X})$ to $(\Sigma_{\forall i}(\omega_i - B_i'), \bar{X}')$ is orthogonal to the production-possibility frontier and the point $(\Sigma_{\forall i}(\omega_i - B_i'), \bar{X}')$ lies on the frontier. Consequently, the adjusted quantity of the public good and the adjusted bid of agent k are

$$\bar{X}' = \frac{\bar{X} + q \sum\limits_{\forall k} B_k}{1+q^2} \tag{9}$$

$$B_k' = \frac{B_k q \bar{X}'}{\sum\limits_{\forall i} B_i}$$

(c) If $\Sigma_{\forall i} B_i < q\bar{X}$, PLATO computes each agent's share of cost $(q - \hat{B}_i/\bar{X}_i)\bar{X}$, which is received as a private message by each i. In this case PLATO bypasses the voting mode, goes to trial $t+1$, and again each agent selects a 2-tuple $(B_i, \bar{X}_i)$ to be transmitted to the center.

(iii) *Stopping rule*:

(a) The process stops on trial $t^* \leqslant T$ if the arrangement resulting from trial t^* is brought to a vote, and all i vote yes.

(b) Otherwise the process stops on trial T.

(iv) *Outcome rule*:

(a) If the process stops by (iiia) the collective has reached an equilibrium agreement, and each agent is paid $V^i[\omega_i - B_i(t^*), \bar{X}(t^*)]$ dollars in cash. If $U_i(\cdot)$ is i's (unobserved) utility of money with $U_i' > 0$, then the final outcome to i from equation (1) in Section II is

$$W_i(t^*) = u^i\left[y_i(t^*), \bar{X}(t^*)\right]$$

$$= U_i\left\{V^i\left[\omega_i - B_i(t^*), \bar{X}(t^*)\right]\right\}$$

(b) Otherwise (i.e., if the process stops by (iiib)) each agent receives $W_i(T) = u^i(\omega_i, 0) = U_i[V^i(\omega_i, 0)]$.

IV. Experimental Design and Results

If subject i has a monotone increasing utility $U_i(V^i)$ for V^i dollars, where $V^i = a_i y^{\alpha_i} X^{\beta_i} - b_i$ is i's payoff function, then the Lindahl equilibrium quantities of the private and public good are independent of the U_i, and given by

$$y_i^0 = \frac{\alpha_i \omega_i}{\alpha_i + \beta_i} \tag{10}$$

$$X^0 = \left(\frac{1}{q}\right) \sum_{\forall i} \left(\frac{\beta_i \omega_i}{\alpha_i + \beta_i}\right)$$

Table 1 exhibits the individual experimental design parameters for the basis set of three classes of subject agents, and Lindahl equilibrium private quantities, bids, and payoffs. The PLATO version of the auction mechanism is designed to replicate the basis economy in multiples of three. That is, there are three classes of individuals identified by endowment and payoff function parameters. An experimental collective consists of k subjects in each class, $k=1,2,3,\ldots$, corresponding to collectives of size $I=3,6,9,\ldots$, in which the unit cost of the public good is $q, 2q, 3q, \ldots$. The public design parameters applying to the experimental collectives and number of experiments conducted under each condition are listed in Table 2. Whenever the available time was adequate, an experimental session consisted of two back-to-back experiments in which the second experiment was a second play with the same "experienced" subjects. The total number of

TABLE 1—EXPERIMENTAL DESIGN, INDIVIDUAL PARAMETERS

Parameter	Parameter Class I	II	III
α	0.2	0.4	0.1
β	0.2	0.1	0.4
a	16	7	14
b	15	6	13
ω	8	10	5
y_i^0, *LE* Private Quantities	4	8	1
$\omega_i - y_i^0$, *LE* Bids	4	2	4
$V^i(y_i^0, X^0)$	\$14.14	\$12.89	\$13.65

Note: *LE* refers to Lindahl equilibrium.

TABLE 2—EXPERIMENTAL DESIGN, COLLECTIVE PARAMETERS

Collective Size I	3	6	9
q, Unit Cost Public Good	2	4	6
X^0, Lindahl Optimal Quantity	5	5	5
X', Free Rider Quantity	2.89	1.67	1.42
Number Experiments:			
Inexperienced Subjects	8	3	4
Experienced Subjects	7	3	4
Total Experiments	15	6	8
Total Subjects	45	36	72

experiments classified by collective size and experience treatment is shown in Table 2. The free rider quantities of the public good listed in Table 2 are computed using (7), which for Cobb-Douglass payoff functions leads to the solution

(11)

$$y_i' = \min\left[\omega_i, \frac{\alpha_i q X'}{\beta_i}\right], X' = \frac{1}{q}\sum_{\forall_i}(\omega_i - y_i')$$

Tables 3, 4, and 5 contain the final trial bids by subject parameter class, and the quantity of the public good provided in each experiment for collectives of size three, six, and nine, respectively. Each experiment is code numbered in the first column of each table. For example in Table 3, PA3.3 refers to PLATO auction, collective size three, session three, in that order. Experiments without a prime designation on the last digit were the first "inexperienced" play while those with a prime were the second "experienced" play. Thus PA3.3 was the first of two experiments in the third session with three subjects, while PA3.3′ was the second experiment using the same subjects. The only difference (besides experience) between two such experiments is that the subjects are rerandomized among the three parameter classes in the second experiment. Experiment PA9.1 aborted due to a program error, but the second experiment in that session PA9.1′ was completed and counted as an experience play. Out of a total of twenty-nine experiments, one collective of size three (PA3.1) and one of size nine (PA9.3) failed to reach agreement to provide units of the public good.

The mean quantity of public good provided by the experimental collectives in each size and experience class is shown in Table 6. Excluded from these calculations are the zero outcomes in the two experimental groups that failed to reach agreement. The pooled overall mean is 5.3 if these two observations are included. From Table 6 it will be seen that there appears to be no systematic effect of collective size or experience on the quantity of public good provided. This is borne out more precisely by a regression of X^* on 1) collective size, 2) a dummy variable for experience, and 3) a dummy variable for the rebate rule. None of these variables are significant: the largest value for the t-statistic for any of these three independent variables is 0.285, with $R^2 = .005$. I must conclude that none of the alleged "treatment" variables has any credible effect on the provision of the public good. Collectives of all sizes provide much larger quantities of the public good than is predicted by the free rider hypothesis, and mean quantities slightly larger than is predicted by the Lindahl quantity. Figure 1 plots all twenty-nine experimental final trial outcomes $(\Sigma_{V_i} y_i^*, X^*)$ along the production-possibility frontier. All twenty-seven experi-

TABLE 3—BID-QUANTITY OUTCOMES, PLATO AUCTION MECHANISM EXPERIMENTS, $I=3$

Subject	1	2	3	X^*, Final Quantity[a]	t^*, Final Trial[c]
Parameter Class	$\omega_1=8$	$\omega_2=10$	$\omega_3=5$		
LE Bid, $w_i-y_i^0=B_{i0}$	4	2	4		
Final Bids, B_i^*					
Rebate Rule 1					
PA3.1[b]	(3)	(1)	(2)	0 (5)	15
PA3.2	4.55	1.8	3.6	5	6
PA3.2′	5	5	4	7	2
PA3.3	2	8	0	5	12
PA3.3′	5	2	1	4	11
PA3.4	3.67	3.67	4.58	6	3
PA3.4′	4.58	6.42	0.92	6	14
Rebate Rule 2					
PA3.5	1.95	3.89	3.89	4.87	12
PA3.5′	3.69	5.54	1.85	5.53	12
PA3.6	3.95	5.92	4.93	7.4	14
PA3.6′	2.94	4.42	4.91	6.13	4
PA3.7	2.85	1.9	2.85	3.8	14
PA3.7′	1	2	3	3	10
PA3.8	3.94	6.89	1.97	6.4	1
PA3.8′	4.89	7.82	1.96	7.33	4

Note: See Table 1.

[a] The *LE* quantity is $X^0=5$, in all experiments.

[b] Bids and quantities in parentheses are the final trial outcomes for experiments failing to reach agreement.

[c] Maximum number of trials, $T=15$.

TABLE 4—BID-QUANTITY OUTCOMES, PLATO AUCTION MECHANISM EXPERIMENTS, $I=6$

Subject	1	2	3	4	5	6	X^* Final Quantity[a]	t^*, Final Trial	T, Maximum Trials
Parameter Class	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$			
LE Bid, $w_i-y_i^0=B_i^0$	4	2	4	4	2	4			
Final Bids, B_i^*									
Rebate Rule 1									
PA6.1	5.76	2.88	2.88	3.83	5.75	2.88	6	3	15
PA6.1′	5.4	3.6	0.9	2.73	4.5	2.73	5	6	15
Rebate Rule 2									
PA6.2	3.92	6.86	0	6.86	8.83	1.9	7.1	8	10
PA6.2′	0	5.99	3.99	3.99	8.98	1.4	6.08	7	10
PA6.3	5.87	6.85	3.91	6.85	3.91	4.89	8.07	8	10
PA6.3′	2.99	4.99	3.99	1	5.98	3.99	5.74	6	10

Note: See Table 1.

[a] See Table 3.

ments reaching agreement produced quantities of the public good in excess of the relevant free rider quantity.

Table 7 records the α (Type I error) and $1-\beta$ (power) probabilities, where the null hypothesis H_0 is that the sample quantities of the public good for each collective size came from a population with mean equal to the theoretical free rider quantities (Table 2), and the alternative hypothesis H_1 is that the sample quantities came from a population with mean equal to the Lindahl quan-

TABLE 5—BID-QUANTITY OUTCOMES, PLATO AUCTION MECHANISM EXPERIMENTS, $I=9$

Subject	1	2	3	4	5	6	7	8	9			
Parameter Class	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$	X^*, Final Quantity[a]	t^*, Final Trial	T, Maximum Trials
LE Bid, $w_i-y_i^0=B_i^0$	4	2	4	4	2	4	4	2	4			
Final Bids, B_i^*												
Rebate Rule 1												
PA9.1′	6.22	3.56	1.78	0.89	7.11	4.44	5.4	2.67	3.56	6	10	15
PA9.2	7.47	2.79	3.71	5.57	1.86	1.76	5.57	8.36	4.64	7	11	15
PA9.3[b]	(1.25)	(6)	(5)	(4)	(1)	(5)	(6)	(7)	(4)	0(7)	(15)	15
PA9.3′	0.97	1.94	3.89	3.89	9.72	1.94	3.89	5.84	3.89	6	14	15
Rebate Rule 2												
PA9.4	3.99	3.99	2.99	4.99	4.99	0	2.99	1	3.99	4.82	8	15
PA9.4′	3	6	1	0	4	1	4	4	2	4.17	12	15
PA9.5	3	5	3	2	1	3	5	7	3	5.33	7	10
PA9.5′	4.99	3.99	0	5.99	4.99	1	5.99	3.99	0	5.16	9	10

Note: See Table 1.
[a] See Table 3.
[b] Bids (quantities) in parenthesis are the final trial bids (quantities) for experiments failing to reach agreement.

tity (5). The first row of selected α probabilities is used to calculate the critical value X_c (for each collective size I), above which we reject H_0 (accept H_1) if the sample mean exceeds X_c. The calculations are based on the t-distribution with sample sizes equal to the number of experiments, and variances computed from the X^*, as listed in Tables 3, 4, and 5. These critical values are then used to compute the $1-\beta$ power probabilities of the test. For example, under H_0 for $I=3$ (col. 1, Table 7), $X_c=2.89+t_\alpha(V_f/N_f)^{1/2}=3.5$ corresponding to a power of $1-\beta=0.7$ (i.e., $t_\beta=(3.5-5)(V_f/N_f)^{1/2}=-0.528$).

Examples of the trial-by-trial "moves" by experimental collectives are illustrated in Figures 2 and 3 for collectives of size three and nine, respectively. In Figure 2 on trial 1, the collective proposed $\bar{X}(1)=8$, and bid $\Sigma_{i=1}^{3} B_i(1)=\Sigma_{i=1}^{2}[\omega_i-y_i(1)]=23-10=13$; on trial 2, $(\bar{X}(2),\Sigma B_i(2))=(8,10)$; and so on. Trials 4 and 11 were voting trials as indicated by the asterisks. On trial 4, the collective overbid the cost of the proposal, i.e., $2\bar{X}=6<\Sigma B_i=8$, and under the integer-increase Rebate Rule 1, the collective's proposal was increased (dashed arrow) to 4. This outcome was vetoed and the collective proceeded to trial 5. Finally, on trial 11 the collective proposed 4, bid 8 and approved

TABLE 6—MEAN PROVISION OF PUBLIC GOOD
$I=3,6,9$; $X^0=5$

	Mean Quantity of Public Good Chosen		
Collective Size I	Inexperienced Collectives	Experienced Collectives	Pooled
3	5.50	5.57	5.53
6	7.06	5.61	6.33
9	5.72	5.33	5.50
Pooled	5.91	5.51	5.7[a]

[a] Based on the twenty-seven "agreement" experiments. The overall mean is 5.3 for all twenty-nine observations.

TABLE 7—ERROR AND POWER PROBABILITIES FOR FREE RIDER (H_0) AND LINDAHL EQUILIBRIUM (H_1) HYPOTHESES

Collective Size	3	6	9
α = Probability of Rejecting H_0 (Accepting H_1) when H_0 is True	0.5	.05	.05
$1-\beta$ = Probability of Accepting H_1 (Rejecting H_0) when H_1 is True	.70	.85	.83
Critical Value X_c	3.5	2.6	2.1

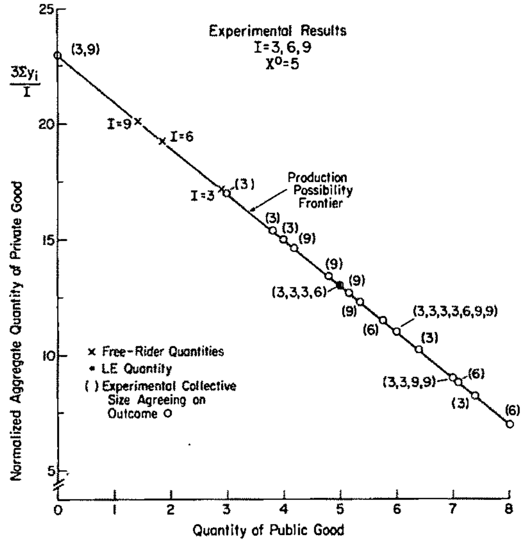


FIGURE 1

the outcome unanimously. Figure 3 illustrates one of the nine-member collectives under the orthogonal rebate rule. Thus on trials 4, 5, and 7, the collective overbid the mean proposal, and in each case a new proposal and bid sum was computed by PLATO based on the orthogonal adjustment (dashed arrows), under Rebate Rule 2, to the boundary of the production-possibility set.

It will be noted that the public good quantity proposals in Figure 3 are less variable than those in Figure 2. This was typical of the larger collectives which, because of this "law of large numbers" effect, were less interactive than the smaller collectives. The variance of final quantities of the public good in three-member collectives was 1.82; in six-member collectives, 1.36; in nine-member collectives, 0.82. This result supports the hypothesis that larger groups are less interactive than smaller groups, but using the Bartlett test, I am unable to reject the null hypothesis that the three population variances are equal.

Since subjects in an auction mechanism experiment are free to bid zero, it is of interest to examine the data of Tables 3–5 for instances of such behavior on the final (equilibrium or boundary) trial. In Table 3,

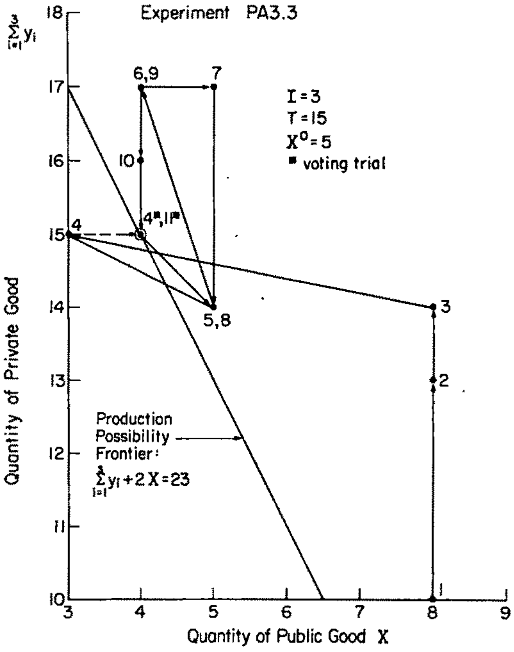


FIGURE 2

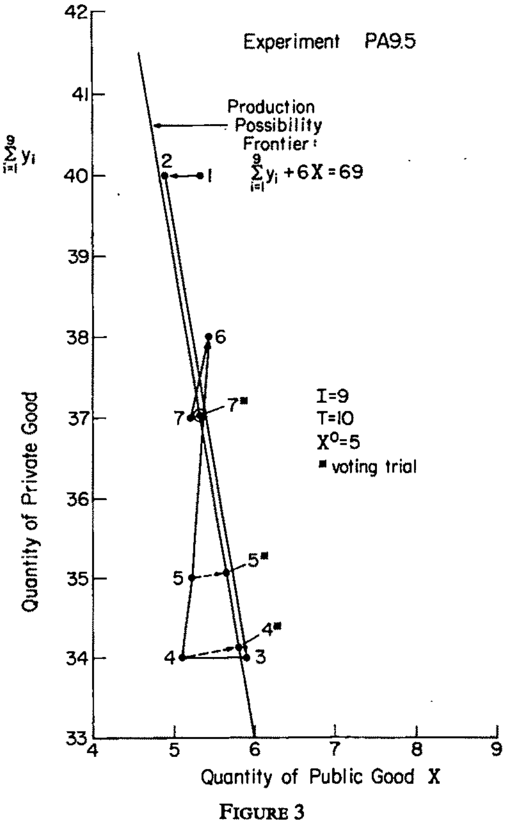


FIGURE 3

TABLE 8—FRACTION OF SUBJECTS BIDDING ZERO OR THEIR ENDOWMENT ON THE FINAL TRIAL

Collective Size	Fraction of Subjects Bidding 0	Fraction of Subjects Bidding ω_i
3	$\frac{1}{45}=.022$	$\frac{3}{45}=.067$
6	$\frac{2}{36}=.056$	$\frac{1}{36}=.028$
9	$\frac{4}{72}=.056$	$\frac{6}{72}=.083$
All Collectives	$\frac{7}{153}=.046$	$\frac{10}{153}=.065$

it is seen that in only one three-member collective did an individual bid zero (subject 3 in PA3.3) on the final trial. At the other extreme, three individuals (subject three in PA3.4, PA3.6, and PA3.6′) bid their entire endowments (before rebate) on the final trial. Normally there was an overbid on a final (agreement) trial with a corresponding rebate that prevented such bids from yielding a zero payoff. Table 8 provides a summary, classified by collective size, of the fraction of subjects bidding zero or their entire endowments. All twenty-nine sessions were included in this determination. If zero bidders represent free rider behavior, and endowment bidders represent anti-free rider behavior, the results in Table 8 suggest that these two extremes are represented about equally in the population of subjects. Furthermore, Table 8 shows no credible increase in free rider bidding behavior as collective size increases in the range from $I=3$–9.

V. On the Pareto Optimality of Experimental Outcomes

The previous section tested experimental public good outcomes against the free rider and Lindahl optimal quantities of the public good. However, as has been emphasized by Samuelson (1969), the Lindahl equilibrium (*LE*) (when there are income effects) provides merely one of many Pareto optimal allocations, and has no special ethical appeal. Since our experimental collectives typically reached agreement on public good quantities that deviated from the Lindahl optimal quantity, it is natural to ask whether these deviations fell within the Pareto optimal set.

The Pareto optimal set is defined by (6). For $I=3$ and Cobb-Douglas utility functions (6) becomes

$$\sum_{i=1}^{3}\frac{\beta_i y_i}{\alpha_i X}=q=\frac{\Sigma(\omega_i-y_i)}{X} \tag{6′}$$

This suggests a linear regression hypothesis among any three of the four quantities (y_1, y_2, y_3, X) depending upon which quantities are arbitrarily selected as the dependent and independent variables. If y_2 and y_3 are regarded as the independent variables with X being the dependent variable, then the Pareto-regression hypothesis is

$$X=\left[\frac{\gamma_2-\gamma_1}{q(1+\gamma_1)}\right]y_2+\left[\frac{\gamma_3-\gamma_1}{q(1+\gamma_1)}\right]y_3+\frac{\gamma_1\sum_{i=1}^{3}\omega_i}{q(1+\gamma_1)} \tag{12}$$

where $\gamma_i=\beta_i/\alpha_i$. Based on the experimental design parameters listed in Table 1, the Pareto-prediction hypothesis is

$$X=-0.1875y_2+0.75y_3+5.75 \tag{12′}$$

Using the fourteen observations on collectives of size three to estimate the coefficients of equation (12′) gives

$$\begin{array}{llll} X= & -0.546y_2 - & 0.497y_3 & +\ 9.533 \\ & (-3.915) & (-9.728) & (5.648) \end{array}$$

$$R^2=.74$$

The *t*-statistics shown in parentheses under each coefficient are computed on the null hypothesis that the population value is given by the corresponding coefficient in (12′). It is clear that these coefficients differ very significantly from their Pareto-predicted values. Hence, we reject the hypothesis that collectives of size three, using the auction

TABLE 9

Parameter Class	I $\omega_1=8$	II $\omega_2=10$	III $\omega_3=5$
LE Bid, $\omega_i-y_i^0$	4	2	4
Mean Bid, All Collectives	3.94	4.86	2.63

Note: See Table 1.

mechanism, tended to provide Pareto optimal allocations.

An examination of mean bids by subjects in the three parameter classes (Table 9) provides some insight into the nature of subject deviations from the Lindahl optimum. Subjects in class I having an endowment of 8, and a theoretical *LE* bid of 4, bid on average very close to 4. However, class II subjects with an endowment of 10 and a theoretical *LE* bid of 2, have a mean bid of 4.86. These subjects much prefer the private good to the public good, but apparently because of their relatively large endowments tend to "overcontribute" to the public good. Similarly, class III subjects with an endowment of 5 and a *LE* bid of 4, tend to "undercontribute" with a mean bid of only 2.63. Hence, the "rich" give more and the "poor" give less than is predicted by the *LE* bids. In this design, such a pattern of deviation is not consistent with Pareto optimality.

These results contrast sharply with my 1979 paper reporting the results of twelve auction mechanism experiments based on designs with no income effects. In these previous experiments, subject i received a payoff table $V_i(X)$ mapping quantities of the common quantity X into private dollars. Subject bids b_i were in dollars per unit of the public good, and the stopping rule required $b_i=q-\Sigma_{\forall j\neq i}b_i$, $\forall i$, on two successive trials with $X_i=\bar{X}$, $\forall i$, on the second trial (see my 1979 paper for a complete discussion). All but one of these partial equilibrium experiments converged to the unique Pareto and Lindahl optimal quantity of the public good given by $\Sigma_{\forall i}V_i'(X^0)=q$. Furthermore, a linear regression of final trial bids b_i^* on the *LE* prices $V_i'(X^0)$ using data on individual subjects across three series of experimental replications yielded intercepts insignificantly different from zero, slopes insignificantly different from one and R^2 values between 0.82 and 0.99. Hence, with no income effects, experimental outcomes are very accurately predicted by the *LE* quantity of the public good and the *LE* price. In the current series of experiments with income effects, the *LE* quantity of the public good is a fair predictor, and the *LE* bids a very poor predictor of experimental outcomes.

VI. Effect of Parameter Changes: How Robust Are the Experimental Results?

It is always appropriate to ask whether particular experimental results are due to artifacts unrelated to the theory one purports to test. In this context, it will be noted that subject payoffs were presented as 12 × 12 tables listing $V^i(y_i,X)$ for the integer values y_i, $X=1,2,\ldots,12$. Hence, the median value of X in a subject's payoff table is 6.5. From Table 6 the pooled mean provision of the public good is 5.7 which is about as close to this median as it is to the theoretical *LE* quantity (5). Also, look at Table 9 listing mean bids. On average subjects contribute approximately one-half their endowments to the public good and retain one-half for private use. With total endowment 23 and $q=2$ in a collective of three members, under the one-half rule the total bid is 11.5, providing budget support for 5.7 units of the public good. In such low-information games as the auction mechanism, it is not unreasonable to conjecture that subjects are drawn to median quantity proposals, or a simple

TABLE 10—EXPERIMENTAL DESIGN: $I=6, q=2, X^0=9, X'=3.33, T=10$

Parameter	Parameter Class I	II	III
α	.24	.96	.8
β	.96	.24	.8
a	1.5	1.5	1.5
b	0	0	0
ω	5	10	6
y_i^0, *LE* Private Quantities	1	8	3
$\omega_i-y_i^0$, *LE* Bids	4	2	3

Note: See Table 1.

TABLE 11—BID-QUANTITY OUTCOMES, PLATO AUCTION MECHANISM EXPERIMENTS; $I=6, q=2, X^0=9, X'=3.33, T=10$

Subject	1	2	3	4	5	6		
Parameter Class	I $\omega_1=5$	II $\omega_2=10$	III $\omega_3=6$	I $\omega_1=5$	II $\omega_2=10$	III $\omega_3=6$	X^*, Final Quantity	t^*, Final Trial
LE Bid, $\omega_i-y_i^0=B_i^0$	4	2	3	4	2	3		
Final Bids, B_i (Rebate Rule 2)								
PA1	3	3	3	2	1	3	7.5	9
PA1′	2.98	4.96	2.98	2.98	0.99	0.99	7.93	10
PA2	(1.5)	(4)	(4)	(0)	(6)	(1)	(8.83)	10
PA3	3.92	3.92	2.94	3.92	2.94	0.98	9.3	7
PA3′	1.97	1.97	1.97	1.97	4.93	1.97	7.4	9
PA4	3.61	4.51	3.61	0.90	4.51	4.51	10.83	1
PA4′	3.81	4.76	4.76	1.90	0	4.76	10	8
PA5	(.99)	(9.88)	(3.95)	(3.95)	(1.98)	(1.98)	(11.37)	10
PA5′	3	1	1	4	10	3	11	10

Note: See Table 1.

TABLE 12—MEAN BIDS AND MEAN PROVISION OF PUBLIC GOOD

Parameter Class	I $\omega_1=5$	II $\omega_2=10$	III $\omega_3=6$
LE Bid $\omega_i-y_i^0=B_i^0$	4	2	3
Mean Bid	2.85	3.46	2.81
Mean Provision of Public Good 9.14			

Note: See Table 1.

50–50 division of endowment between the public and private expenditure alternatives.

These arguments suggest the desirability of replicating the experiments with a different set of parameters. Table 10 lists the parameters used in a second series of nine experiments. For variation on the previous design, this design uses parameters $\alpha_i+\beta_i>1$ yielding only quasi-concave $V^i(y_i, X)$ functions. Two subject classes in the present design retain the same endowments (5 and 10) and *LE* bids (1 and 8) that characterized the previous design.

The final trial bids and quantities of the public good are shown in Table 11. Seven of the nine experimental collectives reached agreement. The mean provision of public good and the mean bids by parameter class are listed in Table 12. A t-test of the null hypothesis that the population mean quantity is the *LE* 9 cannot be rejected ($t_L=0.24$), whereas the free rider hypothesis is rejected ($t_F=10.1$). It is also clear that collectives are not providing quantities of the public good near the artifactual median X in their payoff tables. Comparing the mean bids of class III and class II subjects in Table 9 with the mean bids of class I and class II subjects in Table 12 (endowments 5 and 10, respectively), note a modest increase in the bids of subjects with endowments of 5 and a considerable decrease in the mean bid of subjects with endowments of 10. This is not consistent with the artifactual one-half endowment bidding rule.

I conclude that the auction mechanism results reported in Section IV are robust with respect to parameter changes. Both sets of experimental results support a slight overprovision of public goods relative to the Lindahl optimum. However, individual bids do not support Lindahl optimal bids, with the deviation of actual from Lindahl optimal bids being explained by a consistent tendency for endowment-rich subjects to overcontribute, and endowment-poor subjects to undercontribute, in relation to the Lindahl optimum.

REFERENCES

P. Bohm, "Estimating Demand for Public Goods: An Experiment," *Euro. Econ. Rev.*, No. 2, 1972, *3*, 111–30.

E. Brubaker, "Free Ride, Free Revelation, or Golden Rule?," *J. Law Econ.*, Apr. 1975, *18*, 147–61.

E. Clarke, "Multipart Pricing of Public Goods," *Pub. Choice*, Fall 1971, *2*, 17–33.

______, "A Market Solution to the Public Goods Problem," Urban Econ. Repts., Univ. Chicago, 1968.

J. Ferejohn and Fiorina, "The Paradox of Not Voting: A Decision Theoretic Analysis," *Amer. Polit. Sci. Rev.*, June 1974, *68*, 525–36.

______, **R. Forsythe, and R. Noll,** "An Experimental Analysis of Decision Making Procedures for Discrete Public Goods: A Case Study of a Problem in Institutional Design," in Vernon Smith, ed., *Research in Experimental Economics*, Greenwich 1979.

J. K. Galbraith, "Economics as a System of Belief," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 469–78.

J. Green and J.-J. Laffont, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, Mar. 1977, *45*, 427–38.

T. Groves, "Incentives and Teams," *Econometrica*, July 1973, *41*, 617–33.

______, "The Allocation of Resources Under Uncertainty: The Information and Incentive Roles of Prices and Demands in a Team," tech. rept. no. 1, Center Res. Manage. Sci., Univ. California-Berkeley, Aug. 1969, ch. 4, 71–73.

______ **and J. Ledyard,** "Optimal Allocation of Public Goods: A Solution to the Free-Rider Problem," *Econometrica*, May 1977, *45*, 783–809.

L. Hurwicz, "On Informationally Decentralized Systems," in Roy Radner and C. Bart McGuire, eds., *Decision and Organization*, Amsterdam 1972.

J. Ledyard and J. Roberts, "On the Incentive Problem with Public Goods," disc. paper no. 116, Center Math. Stud. Econ. Manage. Sci., Northwestern Univ. 1974.

G. Marwell and R. Ames, "Experiments on the Provision of Public Goods I: Resources, Interest, Group Size, and the Free Rider Problem," *Amer. J. Sociol.*, May 1979, *84*, 1335–60.

Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups*, New York 1968.

P. Samuelson, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

______, "Pure Theory of Public Expenditure and Taxation," in Julius Margolis and Henri Guitton, eds., *Public Economics*, London 1969.

V. Smith, "Incentive Compatible Experimental Processes for the Provision of Public Goods," in his *Research in Experimental Economics*, Greenwich 1979.

______, "Relevance of Laboratory Experiments to Testing Resource Allocation Theory," in Jan Kmenta, ed., *Evaluation of Econometric Models*, forthcoming.

G. Stigler, "Economic Competition and Political Competition," *Publ. Choice*, Fall 1972, *13*, 91–109.

J. Sweeney, "An Experimental Investigation of the Free-Rider Problem," *Soc. Sci. Res.*, 1973, *2*, 277–92.

Gordon Tullock, *Toward a Mathematics of Politics*, Ann Arbor 1967, ch. 7.

W. Vickrey, "Counterspeculation, Auctions and Competitive Sealed Tenders," *J. Finance*, Mar. 1961, *16*, 8–37.

A Model of the Firm as a Stockholder-Employee Cooperative Game

By MASAHIKO AOKI*

In the traditional neoclassical approach, firms are nothing more than technological black boxes through which inputs of marketed productive factors are transformed into marketable outputs. In this paper, however, I present and analyze a model of the firm which emphasizes its aspect as a quasi-permanent organization of stockholders and employees.

The employees are considered to embody skills and knowledge more or less specific to the firm as a result of quasi-permanent association with it.[1] The employees, in cooperation with physical assets supplied by the stockholders, can produce some economic gains which would not be possible through mere casual combination of marketed factors of production. These economic gains accrue to the firm from the unique and lasting interaction of the organizational resources, both human and physical, and may be termed the organizational rent. Through the acquisition of firm-specific skills and knowledge, the employees may be able to exert implicit or explicit bargaining power over the disposition of the organizational rent. A similar observation was made by Alfred Marshall:

> The point of view of the employer however does not include the whole gain of the business: for there is another part which attaches to his employees. Indeed in some cases and for some purposes, nearly the whole income of a business may be regarded as...a *composite quasi-rent* divisible among the different persons in the business by bargaining, supplemented by custom and by notion of fairness... .[p.626]

Such a viewpoint seems to be increasing in importance as the internal organization of firms become highly extensive in modern capitalistic economies. Firms are not typically the mechanical combination of atomistic agents. The object of this paper is to formalize the nature of the bargaining, and to analyze its outcome and its implication for the firm's price-setting and growth-planning policy.

If the employees of the firm receive nothing more than an exogenously determined wage, then the conventional neoclassical model of the stockholder-controlled firm (hereafter referred to as the *S* firm) would be relevant. On the other hand, if the whole organizational rent is under the control of the workers, the firm may be modelled à la Benjamin Ward and Evsey Domar. Whereas the stockholder-controlled firm maximizes the value of the firm and the worker-controlled firm (the *W* firm) maximizes income per worker, the real firm ought to behave somewhere between these extremes. In-between, however, there are infinitely many ways of distributing internally the organizational rent. A choice of distribution may be made through such diverse institutional arrangements as collective bargaining in the United States and in the United Kingdom, codetermination (workers' participation in management) in West Germany, and conciliatory adjudication by the managers as is the case in Japan. Each mode of distribution would be associated with a particular pattern of price setting and growth policy of the firm.

In Section I below, I make assumptions concerning the market conditions facing the firm. A tradeoff relation between the maximum stock value and employees' income

---

*Professor, Institute of Economics, Kyoto University. The helpful comments and suggestions by Robin Marris, David Starrett, Ryutaro Komiya, George Borts, and an anonymous referee of the *Review* are gratefully acknowledged.

[1] See the vast and growing literature on the internal labor market.

level derived under those assumptions will be called the bargain-possibility frontier. If the firm acts efficiently as an organization, a point on the frontier may be reached. But which point?

In Section II, I specify the nature of the distributive bargaining between the stockholders and the employees as a two-person cooperative game à la Frederik Zeuthen, John Nash, and John Harsanyi, and discuss the determination of an internal distribution on the bargain-possibility frontier. Of course the stockholders do not sit at the bargaining table with the employees to negotiate on the stock value. The former delegate bargaining power to the manager who negotiates with the latter on the wage level. However, in the model presented later, the manager does not act as a mere representative of the stockholder collectivity. His role is to estimate the relative bargaining power between the stockholders and the employees, and to formulate the firm's policy of internal distribution, sales price, and expenditure for sales growth, which would lead to an organizational equilibrium of bargaining powers. A measure of bargaining power of each party is defined in terms of its external opportunity outside the firm and its attitude toward risk. In stressing the autonomous role of the manager, the present model is somewhat akin to the "managerial" model as advanced by Robin Marris and others. But instead of maximizing his own utility, the manager arbitrates between the stockholders and the employees to promote organizational homeostasis. He is an organization man.

In Section III, the equilibrium solution of sales price and expenditure for sales growth will be characterized. It includes optimal policies for the S firm and the W firm as special cases.

I. The Bargain-Possibility Frontier

In this section, I derive the set of efficient combinations of the firm's stock value with the employee's share in the organizational rent. I call it the bargain-possibility set. Its nature depends partly upon market conditions and partly upon the internal structure of the firm.

It is assumed that time extends indefinitely through periods of equal duration. At the beginning of the current period, the firm's sales price is assumed to be fixed for that period. At the same time, "growth expenditure" for sales promotion, new equipment, personnel training, and the like are made. The effects of these expenditures will not be realizable until the next period, when the market demand for the firm's product will expand in response to the firm's sales effort and when its investment in both human and physical capital will become productive. The manager is guided in his price-setting and growth-expenditure decisions by his expectations as to market conditions. His expectation for the current period is called short term and that which covers future periods is called long term.

The manager's short-term expectation is summarized in the subjective demand function for the firm's product:

$$x = \alpha p^{-(1/\eta)} \tag{1}$$

where x is the amount of sales expected in the current period under the sales price p and η the inverse of price elasticity of demand.

The manager's long-term expectation is summarized by what I will call the growth-cost function, defined as follows. Let us assume a simplified situation in which physical equipment is subject to complete depreciation within a period.[2] In this situation a decision about investment in new equipment can be made with regard to a planned growth rate only, independent of the amount of the existing stocks. Then we may assume that there is a simple correspondence between a planned growth rate g of sales from the current period to the next, under a fixed price, and the average growth expenditure T per unit of current sales necessary to realize the growth rate. Thus we have

$$T = \phi(g) \tag{2}$$

---

[2] If the possibility of the fixity of durable equipment is explicitly introduced into the model, an interesting implication for the price-setting behavior peculiar to the firm as a stockholder-employee cooperative game would result. See fn. 11 below.

where $\phi(\cdot)$ is a monotone-increasing, convex function.

Payment to the employees is composed of two parts. The first part is equal to wages that any worker can earn for the same period of time in the competitive external labor market without any quasi-permanent association with organizations. This part is treated as given in the model. The second, or residual, which I will call the *differential earnings* of the employee and which accrues to them in such forms as monopolistic wage, bonus, profit sharing, etc., is due to their quasi-permanent association with the firm and is an endogenous variable of the model.

Let us denote the average constant prime cost of output by c which includes the (competitive) labor cost net of the payment of differential earnings. Define *organizational rent* of the firm π as total revenue net of growth expenditures and the prime cost, i.e.,

$$\pi=[p-c-T]x \tag{3}$$

where it is assumed that all the growth cost is financed internally. This net cash flow is to be divided up among the stockholders and the employees in the form of dividends and differential earnings.

Let θ be the stockholders' share of the organizational rent and l the number of employees necessary for a unit of product. Then average differential earnings ω of the employees is given by

$$\omega=\frac{(1-\theta)\pi}{lx} \qquad (0\leqslant\theta\leqslant 1) \tag{4}$$

Here it is assumed that the amount of employment can be adjusted exactly to the requirement of expected short-term demand at the beginning of the period. However, the employees are supposed to exert effective constraint on the management for the continued employment of the N employees already in the firm. Then the manager is constrained in the choice of sale price by the condition:

$$(N/\alpha l)\leqslant p^{-(1/\eta)} \tag{5}$$

It is assumed that the employees are interested in the amount of their lifetime earnings at the employing firm. (If an outside opportunity looks better, then the employees can of course switch to it.) For the sake of simplicity, economies of scale in production and technological progress are assumed away throughout the paper. Then, as far as the employees' relative share $1-\theta$ is expected to remain unchanged over periods, so is the average differential earnings per employee. At this point, let us introduce a further drastic assumption that the employees are homogeneous and are remunerated equally. In this case, given a discount rate ρ of future earnings and an expected length H of remaining working periods of a representative employee, the present value of his lifetime differential earnings at the firm would be equal to the current differential earnings ω times $[1-(1/(1+\rho))^H]/\rho$. Therefore, a level of well-being of the representative employee may be measured by the amount of current differential earnings ω.[3]

On the other hand, end-of-period returns to the new investors who invest in the stocks of the firm are composed of two parts, dividends and capital gains. In general, the rate of stock appreciation depends upon investors' expectations about those factors which determine the future stream of dividends, such as the state of the firm's organization, managerial policies, and the market conditions. Therefore, there cannot be any simple, satisfactory theory of capital gains. In this paper, I avoid this difficulty by adopting a very simple assumption: investors replace their uncertain knowledge of

---

[3] One may argue, however, that the employees also have a preference for growth of the firm. This would certainly be a possibility, especially if we consider the hierarchical nature of the internal organization of the firm. Within a hierarchical organization, the higher one's ranking in the hierarchical order, the higher are one's differential earnings (see Herbert Simon). Therefore the expected lifetime earnings of the representative employee would be enhanced by increased probability of promotion, which is the case when the firm grows at a faster speed. The possibility is modelled and its implication for the policy of the firm analyzed in the context of a labor-managed firm in my 1978 paper. See also Marris, pp. 89–99.

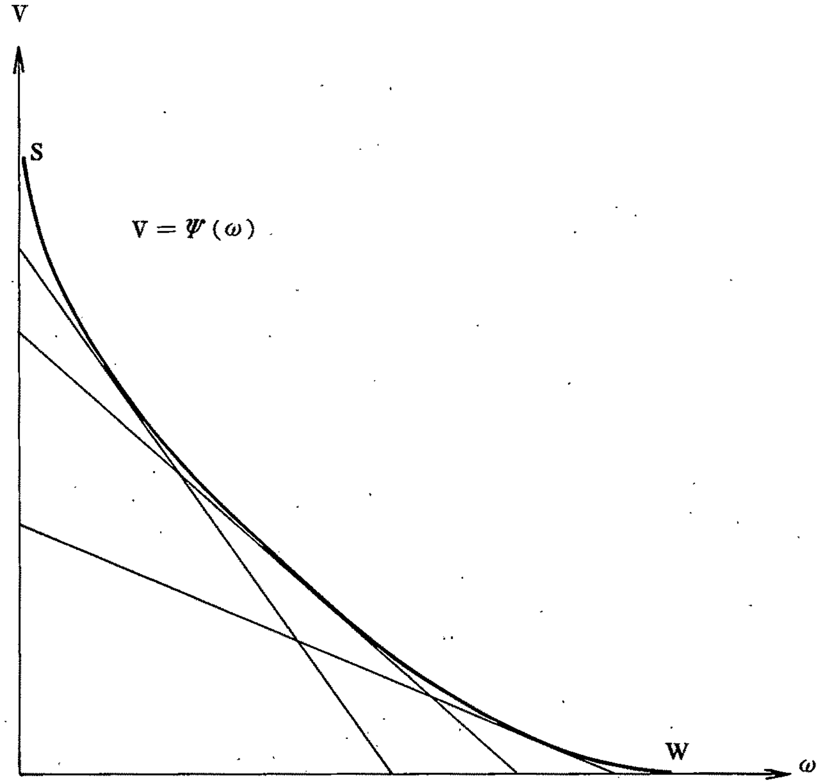


FIGURE 1

future events by a simple Keynesian "convention," that is, they believe on average that the existing state and the firm's policies will continue indefinitely. Then, the rate of stock-value appreciation would be equal to the growth rate currently aimed at by the manager of the firm. The dividends at the end of the current period would equal $\theta\pi$. In the absence of market uncertainty, the competitive valuation V of the firm's stock must satisfy the following relation:

$$gV+\theta\pi=\rho V$$

where ρ is the one-period interest rate. The left-hand side represents the total end-of-period returns to the new investments in the stock of the firm and the right-hand side represents the interest income which investments of the same amount in the one-period bonds would yield. After a suitable arrangement, we can derive[4]

$$V=\frac{\theta\pi}{\rho-g} \tag{6}$$

The "current" stockholders are interested in the maximization of the stock value V.[5]

Summarizing, decision variables of the firm are the sales price p, growth rate g, and stockholders' share θ in the organizational rent π. For each choice of p and g, there will be a given level of π, and all possible allocations between the stockholders and the employees would yield a straight-line trade-off relation between the stock value and differential earning level. By considering all possible choices of p and g subject to the constraint (5), there is a frontier curve as depicted in the Figure 1. Call the curve the *bargain-possiblity frontier* and denote it by

---

[4] In order that the valuation V be finite, the growth rate must be less than the discount rate ρ. I assume that $(N/\alpha l)^{-\eta}=\phi(\bar{g})$ for $\bar{g}$ such that $\bar{g}<\rho$. Then it is never profitable for the firm to plan a growth rate more than $\bar{g}$ and we do not have to be bothered by the possibility of the explosion of the value V.

[5] Here current stockholders are those who hold shares initially before the opening of the stock market in the current period. Roy Radner calls them the *ex ante* stockholders.

$$V=\Psi(\omega) \tag{7}$$

The point S with null employees' differential earnings represents the maximum stock value realized for the firm under the full control of the stockholders. The point W maximizing the employees' average earnings represents the possibility of the worker-managed firm. In-between, we observe possible consequences of "creeping worker's management" of the firm.

The next problem is: which point on the bargain-possibility curve has the most stabilizing effect on the organization of the firm?

II. The Bargaining Game in the Firm

The employees and physical assets supplied by the stockholder collectivity cooperate in the production of organizational rent, even though there is a conflict over its division. This conflict may also extend to the amount of the organizational rent to be produced in each period. For the higher the average growth expenditure, the lower is the current organizational rent per employee. As a result, for any given internal distribution, the lower is the average differential earnings of the employees. In contrast, growth expenditures are necessary costs to the increment of the stockholders' wealth up to a certain amount (see equation (6)). This introduces an important asymmetry into their interests.[6]

Although employees' and stockholder's opinions about the amount and distribution of organizational rent may differ, they cannot injure each other unless they cease to cooperate and thus incur the cost of open conflict. They therefore may have a mutual interest in reaching open and/or implicit agreements before production. This is considered a typical cooperative-type game situation.

Suppose that a preproduction bargaining game for a binding agreement on differential earnings is carried out with the manager of the firm as an arbitrator. The representative employee facing the manager in the bargaining has a von Neumann-Morgenstern (NM) utility indicator $u(\cdot)$ defined on the domain of possible differential earnings. Note that differential earnings are uncertain until an agreement is reached even if there is no market uncertainty. The NM utility indicator represents the employee's attitude toward the risk of internal conflicts in the bargaining.

The investors who became stockholders of the firm through portfolio selection in the financial market in general differ from each other in their attitudes towards risk (and in their expectations). But, needless to say, they neither negotiate with the representative employee individually, nor make decisions on sales price and growth expenditures by themselves. They delegate these jobs to the manager. The manager is assumed to be guided in the bargaining process by a single NM utility indicator $v(\cdot)$ defined on the domain of aggregate wealth V producible by the firm. The utility indicator is supposed to represent the manager's view as to stockholders' interest. The interesting question arises whether the manager's actions can ever be approved unanimously *ex post* by the stockholders when he is guided by the single utility indicator. To put it differently, if the manager's utility indicator, constructed in an appropriate way, is replaced by the true utility indicator of any stockholder, *ceteris paribus*, will the same bargaining outcome emerge?[7] This question will be answered after a description of the nature of bargaining.

---

[6] As indicated in fn. 4, the employees may be interested in the growth of the firm if we introduce the hierarchical nature of the firm into the model. But even so, stockholder and employee preferences for growth would not coincide precisely.

[7] This problem is similar in character to the question of the possibility of stockholders' unanimity under uncertainty when they are different in both their subjective probabilities and their risk attitudes. See Sanford Grossman and Joseph Stiglitz for a handy survey and recent treatment of the theory.

The following formulation of a bargaining process is essentially due to Zeuthen and Harsanyi (1956, 1977).[8] Imagine that at bargaining process time τ, the representative employee contemplates whether or not to make a demand for differential earning h in addition to ω^τ which is an agreed base for bargaining at that time. If he fails in the negotiation, the cooperation of the stockholders via the manager is withheld. In this situation of conflict, his differential earnings would be $\hat{\omega}$, which I treat as given in the model. (The value of $\hat{\omega}$ is zero at worst, which is the case when the employees have no choice other than looking for jobs on the competitive external labor market.)

In order for the demand to be worth making, the expected utility gain $(1-q)[u(\omega^\tau+h)-u(\omega^\tau)]$ must exceed the expected utility cost $q[u(\omega^\tau)-\hat{u}]$ of open conflict, where q is the employee's subjective estimate of the probability that the manager would reject his demand, and $\hat{u}=u(\hat{\omega})$. In other words, the maximum probability per h of open conflict that the representative employee can tolerate is given by

$$\frac{u(\omega^\tau+h)-u(\omega^\tau)}{h[u(\omega^\tau+h)-\hat{u}]}$$

and as $h \to 0$, this tends to $u'(\omega^\tau)/[u(\omega^\tau)-\hat{u}]$. This expression may be considered an appropriate measure of representative employee's bargaining power at time τ,[9] denoted by β_u^τ.

On the other hand, let the stock value in the event the employees withdraw cooperation be $\hat{V}$, which is treated as exogenous to the model. If the manager succeeds in rejecting the employees' demand for additional earnings h at time τ, the stockholders' net gain would be $v(\Psi(\omega^\tau))-v(\Psi(\omega^\tau+h))$, whereas if he fails in negotiation and if the employees withdraw their cooperation at the manager's rejection, the stockholders loss would be $v(\Psi(\omega^\tau+h))-v(\hat{V})$. By reasoning symmetrical to that for the employees' case, the maximum risk per h of open conflict that the stockholders can tolerate in the manager's view is

$$\frac{v(\Psi(\omega^\tau))-v(\Psi(\omega^\tau+h))}{h[v(\Psi(\omega^\tau+h))-\hat{v}]}$$

where $\hat{v}=v(\hat{V})$. As $h \to 0$, this tends to $-v'(\Psi(\omega^\tau))\cdot\Psi'(\omega^\tau)/[v(\Psi(\omega^\tau))-\hat{v}]$. Divide it by $-\Psi'(\omega^\tau)$ and call the quotient a measure of the stockholders' bargaining power assessed by the manager at time τ. Denote it by β_v^τ.

The essence of Zeuthen's construction is to suppose that each party will make a concession to his opponent when he finds that the latter's readiness to bear the risk of open conflict is firmer. In mathematical notation,

$$h \gtreqless 0 \quad \text{according as} \quad \beta_u^\tau \gtreqless -\beta_v^\tau \cdot \Psi'(\omega^\tau)$$

Harsanyi (1977, p.12) refers to the relation as the "Zeuthen Principle." Another way of writing it is

$$\frac{d\omega^\tau}{d\tau}=\beta_u^\tau+\beta_v^\tau\Psi'(\omega^\tau) \tag{8}$$

The process (8) may be considered either as a model of actual collective bargaining, or as a description of the process of managerial calculation designed to bring stability to the organizational state of the firm. In the latter interpretation, (8) may be read as follows: The manager of the firm changes the internal distribution and market policies in the direction of increasing the weighted sum of average differential earnings of the employees and the stock value of the firm, with the respective measures of bargaining power as the parametric weights.

---

[8] Zeuthen assumes that both bargaining partners are risk neutral. Harsanyi (1956, 1977) treats the case where the numbers of players are fixed. The present model is somewhat unique in the sense that the number of players (employees) is variable, although it is a straightforward application of the Harsanyi model in its essential character.

[9] The measure $u'/(u-\hat{u})$ appears often in economic literature. Robert Aumann and Mordecai Kurz call it the measure of boldness. Costas Azariadis finds that the measure is also relevant in the analysis of "implicit" labor contracts.

The process (8) reaches an equilibrium when the right-hand side of (8) becomes zero, in which $u'/(u-\hat{u}) = -v'\Psi'/(v-\hat{v})$. As Harsanyi (1956, 1977) noted, this is equivalent to the condition that the product $(u-\hat{u})(v-\hat{v})$ is maximized on the bargain-possibility frontier.[10] In other words, the policy consisting of sales price p^*, planned growth rate g^*, and stockholders' share θ^*, corresponding to the equilibrium of the process (8), is the Nash bargaining solution. Conversely, the Nash bargaining solution equilibrates the bargaining process as formulated above.

Now the question raised earlier as to the possibility of stockholders' unanimity can be answered. In the process of negotiation, the stockholders' marginal gains from rejecting employee demands for higher earnings must be evaluated relative to total losses they may incur in the organizational disarray due to open conflict with the employees. The manager's evaluation is represented by the measure $\beta_v = v'/(v-\hat{v})$. Such a measure would be different if the utility indicator of an arbitrary stockholder is substituted unless he and the manager have the same utility indicator (i.e., unless they have the same risk attitude). This means that the manager must use his own judgement in arriving at organizational agreements about employees' earnings. He is not an errand boy of the stockholders, but plays a more autonomous role.

III. Equilibrium Distribution and Market Policies of the Firm

In this section, I shall derive the Nash bargaining solution explicitly and examine its properties. The Nash solution is given by the 3-tuple (p^*, g^*, θ^*) which will maximize $[u(\omega)-\hat{u}][v(V)-\hat{v}]$, subject to the conditions defined by (1)–(6) and $0 \leqq \theta \leqq 1$.

Taking the logarithm of the objective function and forming the Lagrangian L, we have

$$L = log\left[u((1-\theta)(p-c-\phi(g))/l) - \hat{u} \right]$$
$$+ log\left[v(\alpha\theta(p-c-\phi(g))p^{-(1/\eta)}/\right.$$
$$\left.(\rho-g)) - \hat{v} \right] + \mu(p_{max}-p) + \lambda(1-\theta)$$

where μ and λ are Lagrange multipliers.

Differentiating the Lagrangian expression with respect to p, g, and θ, and setting each at zero, we have

$$(9)\quad \beta_u(1-\theta)p\eta/l - \beta_v\theta x[(1-\eta)p-c$$
$$-\phi(g)]/(\rho-g) \leqslant 0 \qquad (=0 \quad \text{if} \quad p < p_{max})$$

$$(10)\qquad \beta_u(1-\theta)/l + \beta_v\theta x[(\rho-g)$$
$$-(1/\phi')(p-c-\phi(g))]/(\rho-g)^2 = 0$$

$$(11)\quad \beta_u/l - \beta_v x/(\rho-g) = 0 \qquad (0<\theta<1)$$

where $\beta_u = \dfrac{u'}{u-\hat{u}}$ and $\beta_v = \dfrac{v'}{v-\hat{v}}$

Solving the three equations for p, T, and θ, we can get the equilibrium solution of the bargaining game.

The level of θ^* would in general reflect the relative bargaining power between the stockholder collectivity and the employees. This is seen as follows. Conceptually, the measure of the representative employee's (manager's) bargaining power as defined in Section II comprises two components: his attitude toward risking open conflict in the organization, and his attitude toward small gains. The measure declines as his differential earnings (the stock value) increase. This is so because his inducement to earn small gains declines rapidly as his income (wealth) increases, if he is risk averse. To cancel out the component that measures his attitude toward small gains, let us divide the measure of his bargaining power at time τ by the Arrow-Pratt measure of absolute risk aversion at the corresponding time. The resulting ratio $\xi_u = (u')^2/u''(u-\hat{u})[\xi_v = (v')^2/v''(v-\hat{v})]$ may be interpreted as measuring *pure* bargaining power of the representative employee (of the manager) as Aumann and Kurz indicated in a different context.

---

[10] Since the bargain-possibility frontier is convex to the origin as depicted in Figure 1, it is not correct to say that the organizational equilibrium is characterized as the *maximization* of the weighted sum of average differential earnings of the employees and the stock value of the firm when the respective measures of bargaining power at equilibrium are taken as parametric weights.

If the measure of pure bargaining power of the representative employee is a constant ξ_u for a relevant range of differential earnings, then $u'/u = -\xi_u u''/u'$. (The utility indicator is normalized in such a way that $u(\hat{\omega}) = 0$.) Exponentially integrating, $u = cu'^{-\xi_u}$. Again integrating $log\ u(\omega) \sim [\xi_u/(1+\xi_u)]\ log(\omega - \hat{\omega})$. Similarly if the measure of pure bargaining power of the stockholders perceived by the manager is a constant ξ_v, $log\ v(V) \sim [\xi_v/(1+\xi_v)]\ log(V-\hat{V})$.

If the utility functions are specified as above, equation (11) can be rewritten as

$$(12)\qquad \frac{(1-\theta^*)\pi^* - W^0}{\theta^*\pi^* - v^0} = \frac{\xi_u}{1+\xi_u} \Big/ \frac{\xi_v}{1+\xi_v}$$

where $v^0 = (\rho - \phi(T^*))\hat{V}$ and $W^0 = \hat{\omega} l x^*$; i.e., net gains of the employees, $(1-\theta^*)\pi^* - W^0$, relative to those of the stockholder collectivity, $\theta^*\pi^* - v^0$, resulting from their association with the firm, is proportional to their relative bargaining power measured by $\xi_u/(1+\xi_u)/\xi_v/(1+\xi_v)$. If the stockholder collectivity is treated as risk neutral by the manager, then ξ_v becomes infinite and, in turn, $\xi_u/(1+\xi_u)/\xi_v/(1+\xi_v)$ becomes $\xi_u/(1+\xi_u)$, and the employees still enjoy a surplus over $\hat{\omega}$, provided $\xi_u > 0$.

Next, substituting the equilibrium relation (11) into (9) and (10), we have the following:

$$(A)\qquad p^* = \begin{cases} (N/\alpha l)^{-\eta} & (0 \leqq \theta^* \leqq \eta) \\ \min\left[\dfrac{1}{\theta^* - \eta}(\theta^*(c+\phi(g^*)), \; (N/\alpha L)^{-\eta} \right] & (\eta < \theta^* \leqq 1) \end{cases}$$

$$(B)\qquad \frac{\theta^*(p^* - c - \phi(g^*))}{\rho - g^*} = \phi'(g^*)$$

Formula (A) is the condition of the equality of the marginal organizational revenue with the marginal organizational cost, where the marginal organizational revenue (cost) is defined as the sum of the marginal revenue (cost) weighted by the stockholders' share θ and the marginal per-employee revenue (cost) times the number of employees weighted by their share $1-\theta$. The former may be termed the marginal revenue (cost) from the stockholders' point of view and the latter, the marginal revenue (cost) from the employees' point of view. The marginal revenue from the stockholders' point of view is $\theta(1-\eta)p$ and that from the employees' point of view is $-(1-\theta)\eta p$, hence the marginal organizational revenue is $(\theta-\eta)p$. The marginal cost from the stockholders' point of view is given by $\theta[c+\phi(g)]$ and that from the employees' point of view is null.[11] Therefore, equating the marginal organizational revenue with the marginal organizational cost leads to the formula. In Figure 2, the curve A represents the graph of condition (A).

Condition (B) is remniscent of the Keynesian condition on the marginal efficiency of capital. Let the discount rate that equates the present value of the additional dividend stream resulting from an extra 1 percent of growth to the marginal cost of the growth be called the marginal efficiency

---

[11] When there is fixed durable equipment, the marginal cost from the employees' point of view is given by K_0/x, where K_0 is the value of initial stocks of the durable equipment (see Ward). Therefore if K_0 is large relative to x, i.e., if the rate of utilization of existing capacity is lower, the organizational marginal cost is higher. Hence if we compare two firms which are characterized by the same equilibrium share, the same equilibrium growth rate and the same degree of monopoly in the product markets, the sales price of the firm with larger idle capacity would be higher. However, in actual time of lower-capacity utilization, the demand factor would exert a depressing pressure on sales price as the degree of monopoly would be weakened. The desired growth rate would also be lower, and accordingly expenditure for growth would be lower. Their net effects on the level of sales price have to be judged on the basis of empirical observation. But the possibility that the firm with large capital stocks and a large body of monopolizing labor force maintains a high price in a time of lower demand would not necessarily contradict the Zeuthen principle.

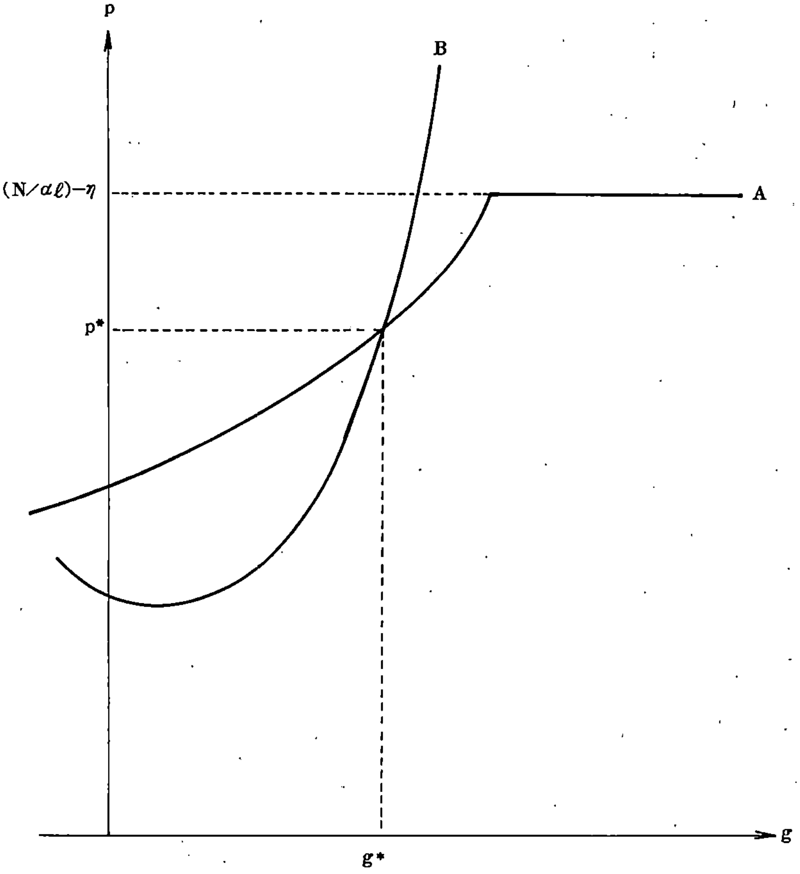


FIGURE 2

of growth. Condition (*B*) states that growth expenditure should be expanded until the marginal efficiency of growth becomes equal to the effective interest rate defined as the difference between the rate of interest ρ and the rate of expected capital gain g. In Figure 2, curve *B* represents the graph of (*B*).

Another way of reading condition (*B*) is possible in terms of James Tobin's "*q*." Multiplying both sides of condition (*B*) by $g/\phi(g)$ and using (6), we have

$$\frac{gV}{\phi(g)x} = \frac{g\phi'(g)}{\phi(g)}$$

The left-hand side is "the ratio of the increment of market valuation to the cost of the associated investment" (Tobin and William Brainard, p. 243), that is, "the *q* ratio on the margin." The right-hand side tells what extra percentage of growth expenditure is necessary to realize a 1 percent increase of sales growth, that is, the elasticity of the growth cost. Its magnitude signifies the relative difficulty for the firm to grow. Condition (*B*) then reads: The growth should be pursued until the marginal *q* ratio becomes equal to the elasticity of growth cost. As the elasticity is greater than one for a positive growth rate by the assumption on the growth cost function, so is the equilibrium value of the marginal *q*.

The equilibrium sales price p^* and the equilibrium planned growth rate g^* are represented by the point where curve *B* intersects with curve *A* from below.

Consider the case where the intersection occurs at $p^* < (N/\alpha l)^{-\eta}$. If the balance of bargaining power is tilted in favor of the representative employee, the equilibrium markup margin would increase and curve A would shift upward. This would have the indirect effect of improving the marginal efficiency of growth. But the decline in the stockholders' share associated with strengthening of the employees' relative bargaining power would have the direct effect of worsening the marginal efficiency of growth, shifting curve B leftward. Both these shifts have the effect of raising the equilibrium sales price. But they have conflicting effects on the planned growth rate. Between them the former effect is proved to be a little stronger (see the Appendix for the proof). If the sales price reaches the maximum $(N/\alpha l)^{-\eta}$, then the further strengthening of the workers' bargaining power entails the slowing down of the planned growth rate.

IV. Concluding Remarks

I have introduced a somewhat unconventional model of the firm in which both the employees and the stockholders share the organizational quasi rent, and where the manager acts as an arbiter between them in formulating corporate policy. The internal distribution of the quasi rent was endogenized in this model together with the sales price and growth expenditure. However, once the internal distribution is given, the equilibrium condition for growth does not differ from the conventional neoclassical condition. This is so because the employees are assumed to be homogeneous and not able to gain from the growth of the firm. An important direction of future research would be to posit that the employees are organized internally into a promotional hierarchy, benefiting from promotion made possible by the growth of the firm, and then examine the implication for equilibrium corporate policy. Such a supposition seems to be more in line with the underlying presumption that the employees are firm-specific resources.

APPENDIX

The sign of $dg/d\theta$:

Differentiating formulae (A) and (B) with respect to θ, we have

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} dp/d\theta \\ dg/d\theta \end{bmatrix} = \begin{bmatrix} \frac{\eta p}{\theta} \\ -r \end{bmatrix}$$

if $p < p_{max}$, where

$$r = p - c - \phi$$

$$m_{11} = -(\theta - \eta)$$

$$m_{12} = \theta\phi'$$

$$m_{21} = \theta$$

$$m_{22} = (1-\theta)\phi' - (\rho - g)\phi''$$

Because the matrix $M = ((m_{ij}))$ is proportional to

$$\begin{bmatrix} L_{pp} & L_{pg} \\ L_{gp} & L_{gg} \end{bmatrix}$$

evaluated at the equilibrium in the direction defined by $L_\theta = 0$, M is negative definite from the maximization of the Lagrangian. Hence $m_{22} < 0$ and $|M| > 0$. Solving for $dg/d\theta$, we have

$$\begin{aligned} dg/d\theta &= -[m_{11}r + m_{21}(\eta p/\theta)]/|M| \\ &= -[-(\theta-\eta)r + \eta p]/|M| \\ &= -\eta r/|M| \leqq 0 \end{aligned}$$

where $dg/d\theta = 0$ only if $\eta = 0$. If $p = p_{max}$, then

$$dg/d\theta = -\phi'(\rho - g)/\theta m_{22} > 0$$

REFERENCES

M. Aoki, "Worker Management and Worker-Owner Sharing in the Hierarchical Work Organization," *Econ. Anal.*

Workers' Manage., forthcoming.

R. J. Aumann and M. Kurz, "Power and Taxes," *Econometrica*, July 1977, *45*, 1137–60.

C. Azariadis, "Implicit Contracts and Underemployment Equilibria," *J. Polit. Econ.* Dec. 1975, *83*, 1183–202.

E. Domar, "The Soviet Collective Farm as a Producer Cooperative," *Amer. Econ. Rev.*, Sept. 1966, *56*, 734–57.

S. Grossman and J. E. Stiglitz, "On Value Maximization and Alernative Objectives of the Firm," *J. Finance*, May 1977, *32*, 389–402.

John C. Harsanyi, "Approaches to the Bargaining Problem before and after the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories," *Econometrica*, Jan. 1956, *24*, 144–57.

______, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge 1977.

Robin Marris, *The Economic Theory of 'Managerial' Capitalism*, New York; London 1964.

Alfred Marshall, *Principles of Economics*, 8th ed., London 1920.

J. F. Nash Jr., "Two-Person Cooperative Games," *Econometrica*, Apr. 1953, *21*, 128–40.

R. Radner, "A Note on Unanimity of Stockholders' Preferences among Alternative Production Plans: A Reformulation of the Ekern-Wilson Model," *Bell J. Econ.*, Spring 1974, *5*, 181–84.

H. Simon, "The Compensation of Executives," *Sociometry*, Mar. 1957, 32–35.

J. Tobin and W. C. Brainard, "Asset Market and the Cost of Market," in *Economic Progress, Private Values and Public Policy: Essays in Honor of William Fellner*, Amsterdam 1977, 235–62.

B. Ward, "The Firm in Illyria: Market Syndicalism", *Amer. Econ. Rev.*, Sept. 1958, *48*, 566–89.

Frederik Zeuthen, *Problems of Monopoly and Economic Warfare*, London, 1930.

Tariffs, Nontraded Goods, and the Optimal Stabilization Policy

By R. F. LUCAS*

The past fifteen years have witnessed a significant expansion of the theory of trade and welfare into the area of distortions. However, as Harry Johnson has noted, this theory is concerned with real distortions such as rigid real wages rather than nominal distortions such as the stickiness of nominal prices and wages. His and subsequent work in this area assumes (either explicitly or implicitly) that monetary and fiscal policies are implemented to deal with nominal distortions thereby introducing a dichotomy into the discussion of the appropriate policies to utilize in combating real and nominal distortions.

The propensity to view nominal distortions as the proper bailiwick of conventional stabilization policies stems from two assumptions. First, it is usually assumed that if there is a nominal distortion, it is a downward sticky nominal wage rate. In this situation, unemployment develops because of too high a real wage. Second, the type of model employed naturally lends itself to a discussion of aggregate disturbances, whether they be of internal or external origin.[1] Consequently, the appropriate stabilization policy consists of expenditure changing which operates on the aggregate price level to generate a real wage consistent with full employment, and expenditure switching which insures balance-of-payments equilibrium.

However, as Robert Barro and Barro and Herschel Grossman (1971, 1976) have demonstrated, if the stickiness of wages is to have a foundation in choice theory, then it logically requires sticky prices as well, and hence the real wage is sticky. In this type of world, firms may operate in the short run where the marginal product of labor differs from the real wage. In addition, if the events of the past five years have demonstrated anything, it is that relative disturbances—that is, disturbances which force a change in relative prices in production sectors—are as important as aggregate disturbances as a source of fluctuations in income and employment. Consequently, in a world of sticky prices and wages and relative disturbances, it would appear that a discussion of the optimal stabilization policy must concern itself with real as well as nominal distortions.

In this paper I analyze the optimal stabilization policy in a world of sticky prices and wages and relative disturbances. A choice-theoretic basis for the investigation is provided by concentrating on the utility level of households rather than on more conventional targets such as full-employment income. The analysis suggests that the optimal policy is one of expenditure switching and expenditure changing coupled with commercial policy which, following the theorem of Jagdish Bhagwati and V. K. Ramaswami, operates directly on the source of the distortion without introducing additional distortions.

Finally, in light of the probability that the optimal policy requires an amount of information which is not available, I examine a second-best policy requiring a minimum

---

*Associate professor of economics, St. Thomas More College, University of Saskatchewan. An earlier version of this paper was presented at the Atlantic Economic Society Conference, October 1977, the American Economic Association Meeting, December 1977 and at the Canadian Economic Association, June 1979. I am indebted to the managing editor of this *Review* for his detailed criticism. I have benefited as well from the comments of John Eckalbar and Don Gilchrist although the usual caveat applies.

[1] There are two general types of models which can be identified. The earlier type consists of two traded goods only one of which is priced in local currency. Examples of this type are Svend Laursen and Lloyd Metzler, Jerome Stein, and Murray Kemp. More recently, models consisting of a composite traded good and a home good have been developed. Two early examples are Robert Mundell (pp. 86–114) and Anne Krueger. In either type, disturbances tend to be characterized as aggregate in nature.

amount of information. One result obtained is the suggestion that tariffs can play a positive role in stabilization policy thereby providing another macro policy tool for the authorities.

I. Analytical Framework

In order to investigate the effects of relative disturbances, I chose a familiar framework in international trade literature—namely, a small, open economy operating under fixed exchange rates which may be parametrically changed. The framework involves six economic goods—labor services, a home good, an import good, an export good, domestic currency, and foreign exchange. The necessity for distinguishing between imports and exports, in contrast to the more common approach which utilizes a composite traded good, follows from the fact that I am interested in the effects of relative disturbances including a change in the world terms of trade between traded goods.

The framework also involves three forms of economic decision-making units—firms, households, and the government. Firms employ labor services to produce the home good and exports. Households supply labor services (exogenously determined), consume the home good and the import good, and save by accumulating money balances. Labor services are homogeneous, freely mobile between the two sectors, and are the only variable input into the production processes. The marginal product of labor is positive but diminishing in both production processes. Households also receive the profits of the firms according to a predetermined pattern and there is no market for claims to these profits.[2]

The home good is not traded internationally because of either high transport costs or an excess level of protection. Domestic currency serves as the medium of exchange, store of value, and unit of account. It is the sole liability of the government.

The prices of the traded goods in terms of foreign currency are fixed. Hence the domestic prices of the traded goods are equal to the exchange rate (defined as the domestic currency price of foreign currency) times the prices in foreign currency. Expectations of exchange rate adjustment are zero and constant.

The model thus includes the following variables:

C = flow of the home good, assumed not traded internationally

X = flow of exports, not consumed at home

V = flow of imports not produced at home

L = flow of labor services where a subscript c or x denotes the labor services assigned to the respective sectors

P_i = the price of the ith good in units of domestic currency; $i=c,x,v$

P_i^* = the price of the ith traded good in units of foreign currency, exogenously determined; $i=x,v$

e = the exchange rate, the domestic currency price of foreign currency; $e \equiv P_x/P_x^* \equiv P_v/P_v^*$

W = the nominal wage rate in units of domestic currency

M = stock of domestic currency

m = rate of change of the stock of domestic currency; $m \equiv dM/dt$

Y = flow of current production, expressed in terms of the home good; $Y \equiv C + (P_x/P_c)X$

In the discussion which follows, I model firm and household behavior under the assumption of general market clearing, and then modify their behavior to take into account the possibility of exchange under non-market-clearing wages and prices.

A. *Firm and Household Behavior under General Market Clearing*

Under general market-clearing conditions, price-taking profit-maximizing firms in each industry set employment and output such that the marginal product of labor in each industry is equal to the nominal wage rate divided by the price of output in that industry. These assumptions are formalized by equations (1)–(4) below:

---

[2] In particular, I assume that profits are only received during the working years. This assumption allows me to specify demand functions as independent of the distribution of income between wages and profits.

$$(1) \quad L_c^d = L_c^d(W/P_c) \qquad L_{c1}^d < 0$$

$$(2) \quad C^s = C^s(W/P_c) \equiv F\left[L_c^d(W/P_c)\right]$$

where $F_{Lc}(L_c) = W/P_c$, and F is the production function in the home-good industry.

$$(3) \quad L_x^d = L_x^d(W/P_x) \qquad L_{x1}^d < 0$$

$$(4) \quad X^s = X^s(W/P_x) \equiv G\left[L_x^d(W/P_x)\right]$$

where $G_{Lx}(L_x) = W/P_x$, and G is the production function in the export industry.

Households are also wage and price takers in all markets, and I assume for simplicity that their supply of labor services is constant. Utility maximization together with the assumption that C, V and future consumption are normal therefore yields the following demand functions.

$$(5) \qquad C^d = C^d(Y, P_v/P_c, M/P_c), \qquad 0 < C_1^d < 1, C_2^d \gtreqless 0, C_3^d > 0$$

$$(6) \qquad V^d = V^d(Y, P_v/P_c, M/P_c), \qquad 0 < (P_v/P_c)V_1^d < 1, V_2^d < 0, V_3^d > 0$$

$$(7) \qquad \frac{m^d}{P_c} = m^d(Y, M/P_c), \quad 0 < m_1^d < 1, m_2^d < 0$$

Equation (7) specifies saving to be independent of the relative price of the import good. I assume that households regard all changes in prices as permanent and that the utility function relating present and future consumption is homothetic. A change in prices alters present and future household constraints in the same manner and hence, such a change should not affect saving measured in either units of the home good or in domestic currency.[3]

The assumption that household expenditure is invariant to changes in the relative price of the import good implies that the degree of substitutability between the home good and the import good is closely related to the size of the relative price elasticity of import demand, denoted by η. Specifically,

$$C_2^d \gtreqless 0 \text{ as } \eta \gtreqless 1,$$

$$\text{where}^4 \qquad \eta \equiv -\frac{P_v/P_c}{V^d} V_2^d$$

The government holds reserves of foreign exchange which it buys or sells in order to fix the exchange rate. This assumed behavior on the part of the authorities allows me to assume balance-of-payments equilibrium even though I am dealing in the short run. The equilibrium does not come about through an intertemporal process but rather through the authorities providing instantaneously the appropriate quantity of financial assets. The government may also alter the money supply through other operations when it engages in a parametric change in the exchange rate.

Combining the behavior of firms, households, and the government, and utilizing the budget constraints to eliminate the flow demand for money yields the conditions which describe general market clearing in a fixed exchange rate regime.[5]

$$(8) \quad C^d(Y, P_v/P_c, M/P_c) = C^s(W/P_c)$$

home-good market

---

[3] For a further discussion of this point, see Kemp and the references cited there.

[4] This implication can be verified by differentiating the household budget constraint with respect to P_v/P_c which yields $C_2^d + (P_v/P_c)V_2^d + V^d \equiv 0$.

[5] Combining the budget constraints of firms, households, and the government yields

$$[C^d(\cdot) - C^s(\cdot)] + (W/P_c)[L_c^d(\cdot) + L_x^d(\cdot) - L^s] + \left[m^d(\cdot) - \frac{dM/dt}{P_c}\right] \equiv 0$$

Since $dM/dt = 0$, $m^d(\cdot)$ must necessarily be equal to zero in a market-clearing situation.

$$dM/dt = P_x X^s(W/P_x) - P_v V^d(Y, P_v/P_c, M/P_c) = 0 \tag{9}$$

foreign exchange market

$$L_c^d(W/P_c) + L_x^d(W/P_x) = L^s \tag{10}$$

labor services market

where $Y \equiv C^s + (P_x/P_c)X^s$, $P_x \equiv eP_x^*$, $P_v \equiv eP_v^*$, $L^s \equiv L^*$. Given predetermined values of L^*, e, P_x^*, and P_v^*, these three conditions determine P_c, M, and W.

B. *Firm and Household Behavior under Non-Market-Clearing Conditions*

Conditions (8)–(10) are only valid if prices and wages adjust rapidly enough to insure that markets always clear. However, if in the short run stickiness of prices and wages prevents markets from clearing, then the above behavior must be modified.

In particular, if P_c and W are temporarily fixed at a level of which implies that $C^d(\cdot) < C^s(\cdot)$, then profit-maximizing behavior implies producing the quantity demanded with the minimum amount of labor services. Thus, the home-good producers demand labor services according to

$$L_c^{d'} = F^{-1}[C^d(\cdot)], F_1^{-1} > 0 \tag{11}$$

where $F^{-1}[\cdot]$ is the inverse of the production function. It follows therefore that if $C^d(\cdot) < C^s(W/P_c)$, the marginal product of labor exceeds the real wage.

Alternatively, if $C^d(\cdot) > C^s(\cdot)$, it is assumed for simplicity that firms will meet demand even if it implies producing at a level of operation where the marginal product of labor is less than the real wage. Consequently, equation (11) depicts home-good producer behavior under conditions of excess demand as well as excess supply.[6]

Export producers never face a sales constraint since by assumption they face a perfectly elastic demand at the world price. Hence their behavior remains as specified above.[7] Also, given that labor supply is constant, household demands have the same form whether or not the labor market is clearing.

Combining the above behavior and again using the budget constraints to eliminate the flow demand for money yields the conditions for determining output and employment under non-market-clearing conditions.

$$C = C^d(Y, P_v/P_c, M/P_c) \lesseqgtr C^s(W/P_c) \tag{12}$$

home-good market

$$dM/dt = P_x X^s(W/P_x) - P_v V^d(Y, P_v/P_c, M/P_c) = 0 \tag{13}$$

foreign exchange market

$$L = L_c^{d'}(C) + L_x^d(W/P_x) \leq L^s \tag{14}$$

labor services market

where $Y \equiv C + (P_x/P_c)X^s$, $P_x \equiv eP_x^*$, $P_v \equiv eP_v^*$. Given predetermined values of e, P_x^*, P_v^*, P_c, and W, these three conditions determine C, M, and L.

It should be noted that system (12)–(14) is an alternative to the system (8)–(10). Equations (8)–(10) may be solved to determine the new general market-clearing position in response to an exogeneous disturbance. However, if P_c and W are sticky, then the equations in conditions (12)–(14) will yield the temporary "disequilibrium equilibrium" generated by the exogeneous disturbance.

---

[6] One might question this assumption since it appears to be incompatible with profit-maximizing behavior. However, the alternative assumption that firms do not meet demand if it exceeds notional supply would require a respecification of import and saving demands to allow for a spillover from unfulfilled home-good demand. This respecification would entail a significant increase in complexity without an offsetting benefit since, as the subsequent analysis demonstrates, the results obtained are invariant to these alternative specifications of producer behavior.

[7] Of course, if there is excess demand in the labor market, then producers in both sectors may be constrained in their labor purchases which would affect supply behavior. However, in the analysis which follows, one target of stabilization policy is clearing in the labor market which rules out the above possibility. For a more general formulation of producer behavior under non-market-clearing conditions, see my dissertation.

II. Stabilization Policy Rankings

In the analysis which follows, I investigate the efficacy of various policy mixes in insulating an economy from relative disturbances and this analysis necessitates a specification of an objective function. Since a country with finite reserves cannot maintain a long-run balance-of-payments deficit, and since it is also subject to a resource constraint, it is assumed that the appropriate criterion for ranking the policy mixes is the maximization of the social welfare function[8]

$$U=U(C,V) \tag{15}$$

subject to

$$P_xX^s(W/P_x)-P_vV^d(Y,P_v/P_c,M/P_c)\leqslant 0 \tag{16}$$

$$L_c^{d'}(C)+L_x^d(W/P_x)\leqslant L^* \tag{17}$$

where U is convex with positive marginal utilities for finite values of C and V.

Ruling out the possibility of corner solutions, a necessary condition for a solution to the above problem is that the marginal rate of transformation between C and V is equal to the marginal rate of substitution between C and V. This condition is depicted in Figure 1 (ignore for the moment the ZZ'' locus); ZZ' represents the production transformation schedule between C and X for the resource constraint L^*. In order to derive a transformation schedule between C and V which incorporates the balance-of-payments constraint as well, simply calculate the amount of imports which can be purchased with the foreign exchange earnings of a given level of export production. For example, for point A, the quantity of imports consistent with balance-of-payments equilibrium is $(P_x^*/P_v^*)X_a$. Utilizing this approach, a transformation schedule between C and V consistent with full employment and balance-of-payments equilibrium can be generated and its slope is given by equation (18).

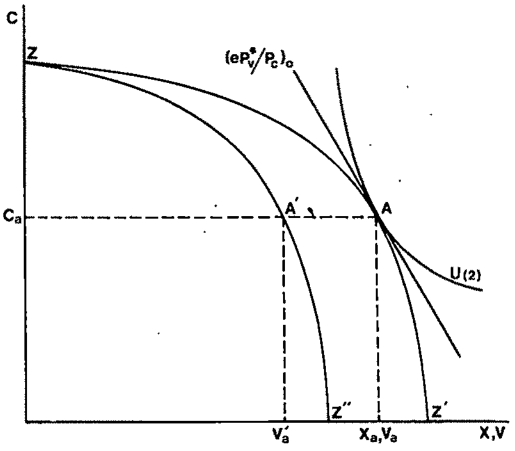


FIGURE 1

$$\frac{dC}{dV}\equiv\frac{dC}{dX}\cdot\frac{dX}{dV}\equiv\frac{dC}{dX}(P_v^*/P_x^*) \tag{18}$$

The position of this transformation schedule is determined by the world terms of trade. By initially choosing units such that $P_v^*/P_x^*=1$, the transformation schedule between C and V must coincide with ZZ'. If $U(2)$ represents a social indifference curve, then point A is the solution to the above problem.

Anticipating the subsequent analysis, it is instructive to consider the impact of a decline in the world price of exports P_x^*. While the production transformation schedule is invariant to changes in P_x^*/P_v^*, the transformation schedule between C and V is not. Since the relative price of imports has risen, the amount of imports which can be purchased with the foreign exchange earnings of a given level of export production must decline. Therefore, for point A, $V_a'=(P_x^*/P_v^*)_1X_a<V_a$ since $(P_x^*/P_v^*)_1<1$. By performing this calculation for all production points on the ZZ' locus, I can trace out a new transformation schedule between C and V (denoted as ZZ'') which must lie interior to ZZ' (except at the production point characterized by complete specialization in the home good). Given the balance-of-payments and resource constraints, the consumption point associated with maximum social welfare must now be bounded from above by the ZZ'' locus.

---

[8] In a more complete analysis, the supply of labor services would enter the utility function opening up the possibility of a policy tradeoff between consumption and leisure. However this consideration is beyond the scope of this paper.

We know that a general market-clearing situation derived under price-taking behavior is isomorphic to the maximum social welfare position. In what follows therefore I begin from a position of general market clearing, impose a relative price disturbance on the system described by equations (8)–(10), and solve for the new general market-clearing position. This then serves as the reference point for the optimal stabilization policy. I then assume sticky prices and wages and solve equations (12)–(14) for the level of output and employment in the absence of discretionary stabilization policy. Finally, I investigate a number of policy mixes which concentrate on maintaining balance-of-payments equilibrium and full employment, demonstrate that they are second or third best policies, and then derive the optimal stabilization policy.

A. *Effects of a Decline in* P_x^*/P_v^*

As an example of a relative disturbance, consider a deterioration in the terms of trade between exports and imports brought about by a decline in the world price of the export good. Within the context of this model, a proportional increase in the foreign currency price of the import good will generate identical general market-clearing solutions.

The decrease in the foreign price of the export good produces a proportional decrease in the domestic price, thereby increasing the real wage in the export industry. Profit maximization requires firms in the export industry to reduce their output. This reduction in exports reinforces the initial reduction in the value of export-good production to generate a balance-of-payments deficit. As the authorities reduce the domestic money supply in the face of this external deficit, the demands for the import and home goods decline through the wealth effect of lower real balances. The resulting excess supply in the home-good and labor markets generates downward pressure on the home-good price and the nominal wage rate.

The new market-clearing position generated by this disturbance is effected by a decline in P_c, M, and W, and it is characterized by a larger, unchanged, or smaller nontraded sector as the relative price elasticity of import demand is greater than, equal to, or less than unity (or alternatively, whether the import and nontraded goods are gross substitutes, independent, or gross complements).[9] Assuming gross substitutability, we can depict the new market-clearing production position in Figure 2 as point B where point A is assumed to be the initial position. As noted above, with a decline in P_x^*/P_v^*, the transformation schedule between C and V shifts inward to ZZ''. Consequently, although production remains on the ZZ' locus, consumption shifts to the ZZ'' locus. The consumption point corresponding to production point B is denoted as B' and it consists of C_b of the home good and $V_b = (P_x^*/P_v^*)_1 X_b$ of the import good. The level of utility is lower in the new market-clearing position since the country has suffered a decline in its external terms of trade. Nevertheless, point B' does represent the point of maximum welfare.

In contrast, if P_c and W are sticky in the short run, then production point B (and hence consumption point B') will not be attained. The reason is straightforward. Recall that the reduction in export production and the subsequent contraction of the domestic money supply generated by the decline in the foreign currency price of the export good creates excess supply in both the home-good market and the labor market. If the home-good price and nominal wage rate fail to respond to this excess supply, firms in the home-good sector have no recourse but to reduce output to the level of sales. This in turn implies a decline in employment in the home-good sector which

[9] See the Appendix for a derivation of these and all other results. For the sake of brevity, the discussion in the text is limited to the case of an elastic import demand although an extension of the analysis to the other cases is straightforward.

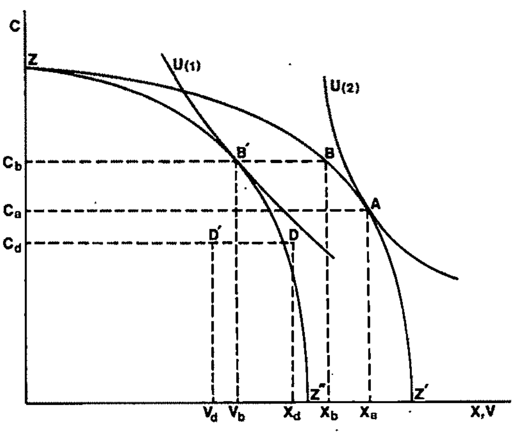


FIGURE 2

reinforces the initial decline in employment in the export-good sector. The resulting reduction in income generates further reductions in demand, output, and employment in the home-good sector through the familiar multiplier process. Thus, in a fixed exchange rate regime with sticky prices and wages, a decline in the world price of exports produces a sharp contraction of output and employment which involves both sectors of the economy.

This result is also depicted in Figure 2. Clearly, production of both C and X must be less than the predisturbance values C_a and X_a, respectively. However, export production must also be less than X_b. To see this, recall that the nominal wage is lower at point B than at point A which implies that under a situation of sticky nominal wages, the real wage in the export sector must be higher than at point B. Hence the production point must lie to the south of A and west of B such as D, with the corresponding consumption point lying to the southwest of B' at D' (with C_d of the home good and $V_d=(P_x^*/P_v^*)_1X_d$ of the import good).

If the government wishes to avoid this sharp contraction in output, it must engage in active stabilization policy. In what follows, I assume that the government objective is to satisfy the equalities in (16) and (17)—namely, balance-of-payments equilibrium and full employment—and demonstrate that this objective must yield a second best solution since it cannot guarantee maximum social welfare.[10]

The first policy mix considered is an exchange rate adjustment and active monetary policy. Differentiating equations (16) and (17) with respect to e, M, and P_x^*, and solving for de/dP_x^* and dM/dP_x^*, yields the changes in e and M necessary to preserve balance of-payments equilibrium and full employment in the face of a decline in the world price of exports—namely, a devaluation of the domestic currency and either a reduction, no change, or an increase in the money supply as the relative price elasticity of import demand is greater than, equal to, or less than unity. This dependence of the appropriate monetary policy on the elasticity of import demand follows from the assumption that the impact on home-good demand of the rise in the domestic price of imports (resulting from the devaluation) is directly related to the elasticity of import demand. In the case of elastic demand, this impact is positive and hence the home-good market is characterized by excess demand following the devaluation, necessitating a contraction of the domestic money supply.[11]

---

[10] This discussion excludes tax-financed government expenditures from the list of available policies since, as Martin Bailey has demonstrated, if these expenditures enter the utility functions of households, relative and aggregate demands are invariant to them unless they are subject to discounting. The utility loss resulting from this discounting implies that tax-financed expenditures must be dominated by other policy tools available to the government. Another alternative is that tax-financed expenditures are productive and as Grossman and I have demonstrated, no discount is required for the provision of these goods to have macro implications. However if the expenditure on these goods is subject to stabilization requirements, there will be a corresponding utility loss to the extent that the provision of these goods deviates from the level judged optimal from the viewpoint of resource allocation.

[11] In the case of unitary elasticity, no active monetary policy is necessary since a devaluation of the domestic currency equiproportionate to the decline in the world price of exports will maintain the government's objective of balance-of-payments equilibrium and full employment. In this special case, this objective will also achieve maximum social welfare since the market-clearing position is unchanged. In this instance, the relative disturbance takes on the properties of an aggregate disturbance which, in the absence of capital account transactions, can be offset entirely by an appropriate exchange-rate adjustment.

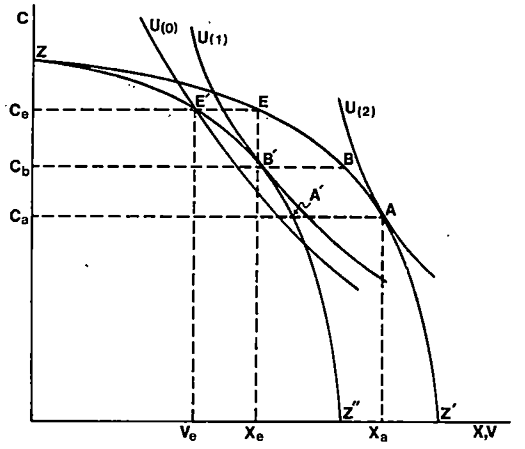


FIGURE 3

This result is depicted in Figure 3 where the original and the new market-clearing production positions are denoted as A and B, respectively. The production position obtained by the above policy mix is denoted by E and the corresponding consumption position by E', and E must lie to the left of B (which of course implies that E' must lie to the left of B').

The proposition that the policy mix of devaluation and contraction of the domestic money supply leads to an overshooting of point B is readily verified. Since the output of the home-good sector has risen in the face of a constant real wage, it follows that the marginal product of labor is less than the real wage. However, in the export sector, the marginal product of labor is equal to the real wage. Consequently, the marginal rate of transformation between C and X must be less in absolute value than the relative price between C and X which requires point E to lie to the left of point B in Figure 3.

This result also implies that, as depicted in Figure 3, the marginal rate of transformation between C and V at point E' must be less in absolute value than the marginal rate of substitution. Formally,

$$(18) \qquad -\frac{dC}{dV} \equiv -\frac{dC}{dX} \cdot \frac{eP_v^*}{eP_x^*}$$

and since

$$(19) \qquad -\frac{dC}{dX} < \frac{eP_x^*}{P_c}$$

it follows that

$$(20) \qquad -\frac{dC}{dV} < \frac{eP_v^*}{P_c} = -\frac{\partial U/\partial V}{\partial U/\partial C}$$

Consequently, even if a policy mix of devaluation and contractionary monetary policy achieves balanced trade and full employment, it cannot achieve maximum utility and consequently, it is a second best policy.[12]

As a second policy mix, consider a combination of exchange rate adjustment and tariff (subsidy) policy.[13] These two policies are independent if the tariff (subsidy) is placed solely on the import good since the tariff (subsidy) changes the domestic terms of trade between traded goods as well as between traded and nontraded goods. Define the ad valorem tariff (subsidy) as t. Then again,

$$P_v \equiv (1+t)eP_v^*$$

---

[12] One might be tempted to infer that this result obtains because my assumption about home-good producer behavior ensures inefficient production. However this is not the case. Suppose home-good producers fail to increase their output levels in response to excess demand. Then the only feasible point on the transformation schedule between C and X is point A which implies a devaluation of the domestic currency that completely offsets the decline in the foreign currency price of X. However, even though in this situation production is efficient, point A', the consumption point, is characterized by inefficiency. This follows from the fact that the policy implies unfulfilled excess demand in the home-good sector since income is constant and P_v/P_c is higher. Consequently, even though the marginal rate of transformation between C and V is equal to the relative price, the marginal rate of substitution between C and V must be less than the relative price. Hence point A' is characterized by inefficiency and again this policy cannot be optimal.

[13] This policy combination was apparently chosen by the Australian government in 1976 when it devalued the Australian currency by 17.5 percent and accompanied the devaluation with a broad reduction in tariffs on imports.

Assume further that the proceeds (expenditure) of the tariff (subsidy) are reimbursed (collected) in a lump sum fashion and that they are allocated to (financed from) consumption and saving. More formally, household disposable income is now given by[14]

$$Y \equiv C + (P_x/P_c)X + T$$

where $$T \equiv t(P_v/P_c)V^d.$$

Substituting these definitions of the domestic price of imports and disposable income into equations (16) and (17), and solving for de/dP_x^* and dt/dP_x^*, yields the changes in e and t necessary to preserve balance-of-payments equilibrium and full employment in the face of a decline in the world price of exports—namely, a devaluation of the domestic currency and either a subsidy or a tariff on the import good as the relative price elasticity of import demand is greater than or less than unity. As in the case of the previous policy mix, under the assumption of elastic import demand, the devaluation generates excess demand in the home-good sector which requires a subsidy on the import-good price in order to eliminate this excess demand through a relative price change.

Furthermore, since the subsidy tends to counteract the impact of devaluation on the value of import demand, while in the previous policy mix the contractionary monetary policy tends to reinforce this impact, it follows that the devaluation necessary to restore balance-of-payments equilibrium is greater when accompanied by a subsidy. This in turn implies that for a given decline in the foreign currency price of the export good, the domestic currency price of this good must be higher under a policy of devaluation and subsidy. Consequently, in Figure 3, the level of export production must exceed X_e or alternatively, the production point must lie to the right of E.

However, it can no longer be assumed that the new position overshoots the new general market-clearing position, and this proposition is most easily demonstrated by concentrating on the ZZ'' locus, the transformation schedule between C and V. By definition, recall from (18) that

$$-\frac{dC}{dV} \equiv -\frac{dC}{dX} \cdot \frac{eP_v^*}{eP_x^*}$$

and since, in an analogous fashion to the first policy mix,

$$-\frac{dC}{dX} < \frac{eP_x^*}{P_c} \tag{21}$$

it follows that

$$-\frac{dC}{dV} < \frac{eP_v^*}{P_c} \tag{22}$$

In addition, however, the subsidy drives a wedge between the world terms of trade P_v^*/P_c^* and the marginal rate of substitution. Therefore, since t is negative,

$$\frac{eP_v^*}{P_c} > \frac{eP_v^*(1+t)}{P_c} = -\frac{\partial U/\partial V}{\partial U/\partial C} \tag{23}$$

Substituting (23) into (22), we obtain

$$-\frac{dC}{dV} \gtreqless -\frac{\partial U/\partial V}{\partial U/\partial C} \tag{24}$$

The inequality in equation (24), when combined with the restriction on the level of export production, implies that the production point generated by a devaluation and tariff can lie anywhere between E and A in

---

[14] It should be noted that this formulation implies that commercial policy affects saving since a rise in t increases the value of income. Consequently, this formulation differs from the standard treatment of commercial policy such as Russell Boyer. However, as Michael Mussa has noted, a tariff on imports will reduce the purchasing power of wealth and it seems reasonable therefore to assume that a tariff will affect saving. I have attempted to capture this effect with the above formulation. In addition, the definition of T assumes that since the impact of a change in t on $(P_v/P_c)V^d$ and hence on the revenue collected is of second-order magnitude, it may be safely ignored.

Figure 3 and the corresponding consumption point between E' and A'.

As I have drawn $U(0)$, any point between E' and A' is preferable to E', but the possibility exists that the exchange rate and tariff policy generates a consumption point where the marginal rate of transformation between C and V exceeds (in absolute value) the marginal rate of substitution by more than the marginal rate of substitution exceeds the marginal rate of transformation in the exchange rate and monetary policy mix. Consequently, it is not possible to demonstrate that either policy is superior although the analysis does suggest that tariffs could play a part in stabilization policy.

By now it should be apparent that, by concentrating on targets of balance of payments and full employment, the government cannot achieve the optimal production point since as long as the real wage in the home-good sector is fixed, the marginal rate of transformation will differ from the marginal rate of substitution.

As Bhagwati and Ramaswami have demonstrated, the optimum policy must operate directly upon the source of the distortion without introducing additional distortions. In the context of this model, the optimum policy is a uniform wage subsidy which achieves the correct real wage in the home-good sector coupled with a devaluation of the domestic currency and a contraction of the domestic money supply. This follows directly from the solution of equations (8)–(10) which, under the assumption of elastic import demand, requires W/P_c to decline, e/P_c to rise, and M/P_c to decline. With constant P_c and W, this can be accomplished by a subsidy on W, devaluation, and reduction in M.

B. *Effects of an Increase in P_x^*/P_v^**

An investigation of an improvement in the terms of trade between exports and imports brought about by (say) a decline in the foreign currency price of imports yields completely symmetrical results. In particular, if import demand is relative price elastic, the new market-clearing position is characterized by a smaller home-good sector and a larger export sector.

With a sticky home-good price and nominal wage, this disturbance generates a multiplicative contraction in income and employment in the home-good sector as producers react to the decline in sales brought about by the relative price change. This decline in income is reinforced by the contraction of the domestic money supply resulting from the balance-of-payments deficit.

A policy mix of devaluation and expansion of the domestic money supply can achieve balance-of-payments equilibrium and full employment through an expansion of the export sector, but again this policy mix generates an overshooting of the new general market-clearing position with a corresponding divergence between the marginal rates of transformation and substitution.[15]

A policy mix of devaluation and tariff on the import-good price can also achieve balance-of-payments equilibrium and full employment with the same ambiguous relationship between the marginal rates of transformation and substitution. Again, the optimal policy must operate on the source of the distortion, the real wage in the home-good sector. In particular, a wage tax, devaluation of the domestic currency, and expansion of the domestic money supply can achieve maximum welfare.

An investigation of additional disturbances, whether they be of external or internal origin, which require a change in the real wage in the home-good sector will generate the same conclusion of the above analysis—namely, that there is in general no second best policy, and the optimal policy must contain an instrument which can operate directly on the real wage in the home-good sector to ensure efficient production while simultaneously ensuring the correct division of demand between imports and the home good and the correct level of aggregate demand.

---

[15] This policy maintains full employment through an expansion of the export sector and a contraction of the home-good sector. If home-good production declines in the face of a constant real wage in this sector, the marginal product of labor must exceed the real wage. However, in the export sector the marginal product of labor is equal to the real wage. Hence, the marginal rate of transformation must exceed, in absolute value, the marginal rate of substitution.

III. Conclusions

The analysis of this paper suggests that, if due to sticky prices and wages, relative demand disturbances are an important source of fluctuations in income and employment, then the study of stabilization policy can be viewed as a subset of the general theory of distortions and welfare. In particular, the optimal policy (i.e., the welfare-maximizing policy) is one which includes a uniform wage tax (subsidy) as well as expenditure switching and expenditure changing instruments.

One should also, however, recognize that the amount of information required by this policy is in all probability in excess of that which is available to the authorities. The rationale for sticky prices and wages would appear to be grounded in the concept of costly information, and therefore it does not seem reasonable to assume that the authorities have perfect information. In a world of costly information, reliance on more conventional targets of stabilization policy such as balance-of-payments equilibrium and full employment is probably reasonable.

The analysis of this paper, therefore, suggests that in a limited information world, tariffs can play a role in stabilization policy. This result follows from the fact that in general the second best policy will generate a level of production where the marginal rate of transformation is unequal to the marginal rate of substitution. Hence, even though a tariff drives a wedge between the marginal rate of transformation and the marginal rate of substitution, it could generate a smaller divergence than other policies.

APPENDIX

The following conditions describe general market clearing in a fixed exchange rate regime:

$$\text{(A1)}\quad C^d(Y, P_v/P_c, M/P_c) - C^s(W/P_c) = 0$$

$$\text{(A2)}\quad dM/dt = P_x X^s(W/P_x) - P_v V^d(Y, P_v/P_c, M/P_c) = 0$$

$$\text{(A3)}\quad L_c^d(W/P_c) + L_x^d(W/P_x) - L^s = 0$$

where $Y \equiv C^s + (P_x/P_c)X^s$, $P_x \equiv eP_x^*$, $P_v \equiv eP_v^*$, $P_v^* = P_x^* = 1$ initially, and $L^s \equiv L^*$. Qualitative assumptions about partial derivatives are $C_1^s < 0$, $X_1^s < 0$, $C_1^d > 0$, $C_2^d \gtrless 0$, $C_3^d > 0$, $V_1^d > 0$, $V_2^d < 0$, $V_3^d > 0$, $L_{c_1}^d = (P_c/W)C_1^s$, and $L_{x_1}^d = (P_x/W)X_1^S$. In addition, $C_1^d + (P_v/P_c)V_1^d < 1$, $C_2^d + (P_v/P_c)V_2^d + V^d = 0$, and in a movement about the product transformation schedule, we consider only the change in Y attributable to the change in P_x/P_c. Also,

$$\xi = -\frac{(W/P_x)}{X^s} X_1^s$$

and

$$\eta = -\frac{P_v/P_c}{V^d} V_2^d$$

Differentiating this system yields the following matrix equation.

$$\text{(A4)}\quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} d(W/P_c) \\ d(e/P_c) \\ d(M/P_c) \end{bmatrix} = \begin{bmatrix} -\varepsilon_1 \\ -\varepsilon_2 \\ -\varepsilon_3 \end{bmatrix}$$

where $a_{11} = -C_1^s$, $a_{12} = X^s(C_1^d + \eta - 1)$, $a_{13} = C_3^d$, $a_{21} = P_c X_1^s$, $a_{22} = -P_c[(W/e)X_1^s + (P_v/P_c)V_2^d + (P_v/P_c)X^s V_1^d]$, $a_{23} = -P_v V_3^d$, $a_{31} = (P_c/W)(C_1^s + X_1^s)$, $a_{32} = -(P_c/e)X_1^s$, $a_{33} = 0$, and where the ε_i are the partial derivatives of the respective market-clearing conditions with respect to the parameter change under investigation. The expression for the above determinant, denoted by Δ, is

$$\Delta = P_c\left[C_3^d + (P_v/P_c)V_3^d\right]\left[(P_c/e)C_1^s X_1^s - (P_c/W)(C_1^s + X_1^s)(X^s)(C_1^d + \eta - 1)\right] - (P_c/W)(C_1^s + X_1^s)(1 - C_1^d - (P_v/P_c)V_1^d) \times (X^s C_3^d) > 0 \qquad \text{if } (C_1^d + \eta - 1) > 0$$

The following conditions determine output and employment under non-market-clearing conditions and in the absence of discretionary stabilization policy.

$$\text{(A5)} \quad C = C^d(Y, P_v/P_c, M/P_c) \gtrless C^s(W/P_c)$$

$$\text{(A6)} \quad dM/dt = P_x X^s(W/P_x) - P_v V^d(Y, P_v/P_c, M/P_c) = 0$$

$$\text{(A7)} \quad L = L_c^{d'}(C) + L_x^d(W/P_x) \leqslant L^s$$

where $Y \equiv C + (P_x/P_c)X^s$, $P_x \equiv eP_x^*$, $P_v \equiv eP_v^*$, $L^s \equiv L^*$.

Differentiating this system yields the following matrix equation:

$$\text{(A8)} \quad \begin{bmatrix} 1 - C_1^d & -(1/P_c)C_3^d & 0 \\ -P_v V_1^d & -(P_v/P_c)V_3^d & 0 \\ -P_c/W & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} dC \\ dM \\ dL \end{bmatrix} = \begin{bmatrix} -\varepsilon_1 \\ -\varepsilon_2 \\ -\varepsilon_3 \end{bmatrix}$$

The value of the above determinant, denoted by Δ', is $\Delta' = (P_v/P_c)V_3^d > 0$.

The following conditions describe the targets of balance-of-payments equilibrium and full employment for government stabilization policy.

$$\text{(A9)} \quad P_x X^s(W/P_x) - P_v V^d(Y, P_v/P_c, M/P_c) = 0$$

$$\text{(A10)} \quad L_c^{d'}(C^d) + L_x^d(W/P_x) - L_s = 0$$

where $Y \equiv C + (P_x/P_c)X^s$, $P_x \equiv eP_x^*$, $P_v \equiv eP_v^*$, $L^s \equiv L^*$.

For the policy of exchange rate adjustment and monetary policy, the relevant matrix equation is

$$\text{(A11)} \quad \begin{bmatrix} X^s(\xi + \eta - (P_v/P_c)V_1^d) & -(P_v/P_c)V_3^d \\ (1/W)X^s(\xi + \eta + C_1^d - 1) & (1/W)C_e^d \end{bmatrix} \cdot \begin{bmatrix} de \\ dM \end{bmatrix} = \begin{bmatrix} -\varepsilon_1 \\ -\varepsilon_2 \end{bmatrix}$$

The determinant of this matrix, denoted by Δ'', is

$$\Delta'' = \frac{X^s}{W}\left[C_3^d(\xi + \eta) + (\xi + \eta - 1)(P_v/P_c)V_3^d\right] > 0 \qquad \text{if } (\xi + \eta - 1) > 0$$

For the policy of exchange rate adjustment and tariff (subsidy) the relevant matrix equation is

$$\text{(A12)} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} de \\ dt \end{bmatrix} = \begin{bmatrix} -\varepsilon_1 \\ -\varepsilon_2 \end{bmatrix}$$

where $a_{11} = X^s(\xi + \eta - (P_v/P_c)V_1^d)$, $a_{12} = X^s(\eta - (P_v/P_c)V_1^d - 1)$, $a_{21} = (1/W)X^s(\xi + \eta + C_1^d - 1)$, $a_{22} = (1/W)X^s(\eta + C_1^d - 1)$. The determinant of this matrix, denoted by Δ''', is

$$\Delta''' = (1/W)(X^s)^2\left[\xi(C_1^d + (P_v/P_c)V_1^d) + (C_1^d + \eta - 1)\right] > 0, \qquad \text{if } (C_1^d + \eta - 1) > 0$$

A. *Change in P_x^**

Differentiating equations (A1)–(A3) with respect to P_x^*, substituting into equation (A4), and solving yields the changes in W/P_C, e/P_c, and M/P_c necessary to preserve general market clearing in the face of a change in the world price of exports.

$$\frac{d(W/P_c)}{dP_x^*} = \frac{-1}{\Delta}\left[P_c X^s(\eta-1)(X_1^s)\right.$$

$$\left.\left(C_3^d+(P_v/P_c)V_3^d\right)\right] \gtreqless 0 \qquad \text{as } \eta \gtreqless 1,$$

$$\frac{d(e/P_c)}{dP_x^*} =$$

$$\frac{-e}{P_c} - \frac{1}{\Delta}\left[(eP_c/W)(C_1^s+X_1^s)(X^s)\right.$$

$$\left.(\eta-1)\left(C_3^d+(P_v/P_c)V_3^d\right)\right] < 0$$

$$\frac{d(M/P_c)}{dP_x^*} =$$

$$\frac{-1}{\Delta}\left[(P_c e/W)\left(1-C_1^d-(P_v/P_c)V_1^d\right)\right.$$

$$\left.(X^s)(\eta-1)(C_1^s+X_1^s)\right] \gtreqless 0 \qquad \text{as } \eta \gtreqless 1$$

Differentiating equations (A5)–(A7) with respect to P_x^*, substituting into equation (A8), and solving yields the changes in C and L generated by a change in the world price of exports under the assumption of a constant home-good price and nominal wage rate.

$$\frac{dC}{dP_x^*} = \frac{(eX^s)(C_3^d)(\xi+1)}{P_v V_3^d} > 0$$

$$\frac{dL}{dP_x^*} =$$

$$\frac{(eX^s/W)\left[\xi\left(C_3^d+(P_v/P_c)V_3^d\right)+C_3^d\right)\right]}{(P_v/P_c)V_3^d} > 0$$

Differentiating equations (A9) and (A10) with respect to P_x^*, substituting into equation (A11), and solving yields the changes in e and M necessary to preserve balance-of-payments equilibrium and full employment in the face of a change in the world price of exports with a constant home-good price and nominal wage rate.

$$\frac{de}{dP_x^*} =$$

$$\frac{-(e/W)X^s}{\Delta''}\left[\xi\left(C_3^d+(P_v/P_c)V_3^d+C_3^d\right] < 0\right.$$

$$\frac{dM}{dP_x^*} = \frac{(e/W)(X^s)^2}{\Delta''}$$

$$\left[(\eta-1)\left(1-C_1^d-(P_v/P_c)V_1^d\right] \gtrless 0 \quad \text{as } \eta \gtrless 1\right.$$

Differentiating equations (A9) and (A10) with respect to P_x^*, substituting into equation (A12), and solving yields the changes in e and t necessary to preserve balance-of-payments equilibrium and full employment in the face of a change in the world price of exports with a constant home-good price and nominal wage rate.

$$\frac{de}{dP_x^*} =$$

$$\frac{-(e/W)(X^s)^2}{\Delta'''}\left[(\xi+C_1^d)\left(C_1^d+(P_v/P_c)V_1^d\right)\right.$$

$$\left.+\left(C_1^d+\eta-1\right)\left(1-C_1^d-(P_v/P_c)V_1^d\right)\right] < 0,$$

$$\text{if } \left(C_1^d+\eta-1\right) > 0$$

$$\frac{dt}{dP_x^*} = \frac{(1/W)(X^s)^2}{\Delta'''}\left[(\eta-1)\left(1-C_1^d\right.\right.$$

$$\left.\left.-(P_v/P_c)V_1^d\right)\right] \gtreqless 0 \qquad \text{as } \eta \gtreqless 1$$

Comparing the change in the exchange rate in the two policy mixes yields

$$\left|\frac{de}{dP_x^*}\bigg|_{dM=0}\right| - \left|\frac{de}{dP_x^*}\bigg|_{dt=0}\right| =$$

$$(\eta-1)\left[C_1^d(P_v/P_c)V_3^d+(\eta-1)\left(C_3^d(1-C_1^d)\right.\right.$$

$$\left.\left.+m_1^d((P_v/P_c)V_3^d)\right)\right] > 0 \qquad \text{if } \eta > 1$$

B. *Change in P_v^**

Differentiating equations (A1)–(A3) with respect to P_v^*, substituting into equation (A4) and solving yields

$$\frac{d(W/P_c)}{dP_v^*}=\frac{1}{\Delta}\Big[P_cX^s(\eta-1)(X_1^s)\times\big(C_3^d+(P_v/P_c)V_3^d\big)\Big]=-\frac{d(W/P_c)}{dP_x^*}\gtreqless 0 \qquad \text{as } \eta\gtreqless 1.$$

$$\frac{d(e/P_c)}{dP_v^*}=\frac{1}{\Delta}\Big[(P_ce/W)X^s(\eta-1)(C_1^s+X_1^s)\cdot\big(C_3^d+(P_v/P_c)V_3^d\big)\Big]\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

$$\frac{d(M/P_c)}{dP_v^*}=\frac{1}{\Delta}\Big[(P_ce/W)\big(1-C_1^d-(P_v/P_c)V_1^d\big)X^s(\eta-1)(C_1^s+X_1^s)\Big]=-\frac{d(M/P_c)}{dP_x^*}0 \qquad \text{as } \eta\gtreqless 1$$

Differentiating equations (A5)–(A7) with respect to P_v^*, substituting into equation (A8), and solving yields

$$\frac{dC}{dP_v^*}=\frac{eX^s(\eta-1)\big(C_3^d+(P_v/P_c)V_3^d\big)}{P_vV_3^d}\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

$$\frac{dL}{dP_v^*}=\frac{(e/W)X^s(\eta-1)\big(C_3^d+(P_v/P_c)V_3^d\big)}{(P_v/P_c)V_3^d}\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

Differentiating equations (A9) and (A10) with respect to P_v^*, substituting into equation (A11) and solving yields

$$\frac{de}{dP_v^*}=\frac{(e/W)X^s}{\Delta''}\cdot\Big[(1-\eta)\big(C_3^d+(P_v/P_c)V_3^d\big)\Big]\gtreqless 0 \qquad \text{as } \eta\gtrless 1$$

$$\frac{dM}{dP_v^*}=\frac{(e/W)(X^s)^2}{\Delta''}\cdot\Big[(1-\eta)\big(1-C_1^d-(P_v/P_c)V_1^d\big)\Big]\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

Differentiating equations (A9) and (A10) with respect to P_v^*, substituting into equation (A12) and solving yields

$$\frac{de}{dP_v^*}=\frac{(e/W)(X^s)^2}{\Delta'''}\cdot\Big[(1-\eta)\big(C_1^d+(P_v/P_c)V_1^d\big)\Big]\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

$$\frac{dt}{dP_v^*}=\frac{(1/W)(X^s)^2}{\Delta'''}\cdot\Big[(1-\eta)\big(1-C_1^d-(P_v/P_c)V_1^d\big)\Big]\gtreqless 0 \qquad \text{as } \eta\gtreqless 1$$

REFERENCES

Martin J. Bailey, *National Income and the Price Level: A Study in Macro-Economic Theory*, New York 1962.

Robert J. Barro, "A Theory of Monopolistic Price Adjustment," *Rev. Econ. Stud.*, Jan. 1972, *39*, 17–26.

______ **and Herschel I. Grossman,** "A General Disequilibrium Model of Income and Employment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 82–93.

______ **and** ______, *Money, Employment and Inflation*, New York 1976.

J. Bhagwati and V. K. Ramaswami, "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy," *J. Polit. Econ.*, Feb. 1963, *71*, 44–50.

R. Boyer, "Commercial Policy Under Alternative Exchange Rate Regimes," *Can. J. Econ.*, May 1977, *10*, 218–32.

H. I. Grossman and R. F. Lucas, "The Macro-Economic Effects of Productive Public Expenditures," *Manchester Sch. Econ. Soc. Stud.*, June 1974, *42*, 162–70.

H. G. Johnson, "Optimal Trade Intervention in the Presence of Domestic Distortion," in Richard E. Caves et al., eds., *Trade, Growth and the Balance of Payments*, Chicago 1965, 3–34.

Murray C. Kemp, *The Pure Theory of International Trade*, Englewood Cliffs 1964.

A. O. Krueger, "The Role of Home Goods and Money in Exchange Rate Adjustments," in Willy Sellekaerts, ed., *International Trade and Finance: Essays in Honour of Jan Tinbergen*, White Plains 1974.

S. Laursen and L. Metzler, "Flexible Exchange Rates and the Theory of Employment," *Rev. Econ. Statist.*, Nov. 1950, *32*, 281–99.

R. F. Lucas, "Monetary Union: An Alternative to Free Exchange Rates," unpublished doctoral dissertation, Brown Univ. 1975.

Robert A. Mundell, *Monetary Theory: Inflation, Interest and Growth in the World Economy*, Pacific Palisades 1971.

M. Mussa, "A Monetary Approach to Balance of Payments Analysis," *J. Money, Credit, Banking*, Aug. 1974, *6*, 333–51.

J. Stein, "The Optimal Foreign Exchange Market," *Amer. Econ. Rev.*, June 1963, *53*, 384–402.

Unemployment, the Allocation of Labor, and Optimal Government Intervention

By DONALD O. PARSONS*

The rate of dispersion of laid-off workers into other jobs in the economy is obviously an important characteristic of the labor market. Labor economists have examined one dimension of this phenomenon at considerable length, namely the effect of unemployment compensation on the intensity of worker search and on the duration of unemployment.[1] In this paper I explore a different dimension of the same process, the effects of industry and economy-wide unemployment rates on the speed of dispersion of unemployed workers into other sectors of the economy.

The rate of intersectoral labor movements has particularly concerned international trade policymakers and theorists. If intersectoral wage adjustments are slow (perhaps due to contracting costs),[2] the market-determined rate of unemployment may not be optimal following a shift in terms of trade,[3] and the subsequent wealth redistribution within the economy may be unnecessarily large.[4] Harry Johnson and Stephen Magee demonstrate that, in the polar case in which factors are completely immobile (and again wages unresponsive), a shift in the terms of trade will generate unemployment in the declining sector with no offsetting gain in production elsewhere in the economy. Government intervention, if costless, may then be justified to restore full employment, perhaps through wage subsidies to firms in the contracting industries.[5]

Lapan argues that government intervention through a variable employment subsidy may be justified to achieve an optimal balance of short-run unemployment costs and long-term allocative gains, even if factors are not immobile. The rate of transfer of unemployed workers to other sectors of the economy and especially the sensitivity of the transfer rate to unemployment levels play a central role in Lapan's analysis.

---

*Professor of economics and research associate, Center for Human Resource Research, Ohio State University. Support for this project was provided by the Bureau of International Labor Affairs, U.S. Department of Labor, and by the Ohio State University Computer Center. Discussions with Harry Grubert and Edward Ray at a preliminary stage were valuable. The comments of Frank Brechling, Harry Gilman, Howard Marvel, and members of workshops at the U.S. Department of Labor and Ohio State University on an earlier draft of this paper are gratefully acknowledged. Thomas Plewes, Bureau of Labor Statistics, kindly provided unpublished data upon which much of this analysis is based. Stephen Cook provided excellent research assistance.

[1] For some recent empirical evidence on the duration of unemployment, see Hirschel Kasper; Hyman Kaitz; George Perry; Arlene Holen and Stanley Horowitz; Ronald Schmidt; William Barnes; John Barron; Ronald Ehrenberg and Ronald Oaxaca. Theoretical aspects of the impact of the unemployment insurance system are considered in Martin Feldstein (1975, 1976), Frank Brechling, and Martin Baily (1974b, 1975).

[2] Arguments for the optimality of wage inflexibility in a world of implicit contracts and worker risk aversion have been made by a number of economists. Insurance-motivated workers may effectively demand wage stability even though the short-run allocative function of wage rates is hindered. See particularly Costas Azariadis; Donald Gordon; Baily (1974a); Herschel Grossman. See also Stephen McCafferty for an interesting uncertainty argument of a different sort which may explain why wage rates are not used as an allocative mechanism. In the case at hand, however, it is transparent that recontracting could make both parties better off *if* recontracting were costless.

[3] Unemployment may be generated (even if wages are completely flexible) if information on alternatives is incomplete. See the contribution in Edmund Phelps' volume, particularly the contribution by Armen Alchian, for an extensive discussion of voluntary unemployed search in response to wage reductions.

[4] This concern underlies the trade adjustment assistance program of the 1974 Trade Act.

[5] John Harris and Michael Todaro and Harvey Lapan discuss alternative policy choices at length. Obviously the subsidy mechanism raises the issue of how the subsidy is financed since most methods of government finance induce economic distortions. Assume here that the subsidies are funded by a per capita tax to which fertility and mortality are insensitive. Alternatively the government can be viewed as mediating (costlessly) the recontracting between worker and firm, in which case the subsidy would be financed by a tax on workers in the declining sector.

Government intervention to alter the *timing* of the release of excess workers and therefore the timing of unemployment in a contracting industry may in fact be optimal in two circumstances, each of which depends ultimately on the structure of the labor transfer function. If increases in unemployment among experienced workers in an industry lead to congestion in local labor markets and a lower absorption rate (the Lapan model), spreading the release of workers from that industry over a period of time may be productive. Moreover, if high levels of unemployment throughout the economy slow absorption of the unemployed workers, postponement of the release of workers until labor market conditions improve may also be optimal.

In Section I below, I develop the optimal intervention arguments of the preceding paragraph more formally. Several critical parameters of the transfer function are identified in that analysis. In Section II, the parameters of the transfer function are estimated using monthly unpublished experienced-worker unemployment data from the Bureau of Labor Statistics (BLS) as well as employee turnover data from published sources over the period 1968–76. The empirical analysis suggests that cyclical smoothing may be optimal, but that phasing the release of workers to avoid congestion in local labor markets apparently is not.

I. An Optimal Intervention Model

A. *The Congestion Model*

The analysis below of optimal labor transfers owes much to the excellent work of Lapan. The congestion cost model, however, is comparable in structure to models of optimal firm investment in physical assets and can be conceptualized in that light.[6] If units of investment become more expensive to acquire and install as the volume of investment rises in any time interval, the profit-maximizing firm will spread profitable investment purchases out over time. Similarly in the labor market model, if unemployed workers face increasing difficulty in obtaining alternative jobs as the number of unemployed workers in the industry rises, an omniscient and costless government planner will find it optimal to slow the release of workers into the unemployment pool from the contracting industry.

To be more concrete, consider the following model. Industry x is a small, competitive industry in the economy, initially in equilibrium. Workers are employed in industry x, label them N_x, to the point at which value of marginal product is equal to the income paid workers, which in turn is equal to the value of marginal product of workers in the economy, or

$$PF'(N_x)=I=v \tag{1}$$

where $P \equiv$ product price of x
$F \equiv$ production function
$I \equiv$ income of workers
$v \equiv$ value of marginal product in the economy as a whole.

The total experienced worker labor force in $x(L_x)$ is the sum of employed and unemployed workers (U), so

$$L_x=N_x+U \tag{2}$$

and the unemployment rate (u) is defined as

$$u \equiv U/L_x \tag{3}$$

Initially, of course, $u=0$ since the industry is in equilibrium and no transfers between firms are required.

Imagine, however, a sudden drop in the price of product x, say the result of an increase in imports of x. If wage rates are not flexible, the profit-maximizing firm will adjust by laying off workers until the equality of value of marginal product and worker income is again reestablished. The displaced workers will presumably find work in other sectors of the economy, including the export sector, after some period of unemployment.

---

[6] A reasonably elementary review of the adjustment cost literature can be found in Brechling. The model discussed below is most closely related to the internal adjustment costs models; see Arthur Treadway.

To the optimal planner, the precise structure of the labor transfer function is crucial. The transfer function can be represented as

$$(4)\qquad dL/dt=-dL_x/dt=\phi(u)U$$

where $\phi(u)$ represents the rate of flow of unemployed workers into other sectors of the economy. If $\phi'=0$, the rate of flow is independent of the unemployment rate among experienced workers in the industry, in which case

$$(5)\qquad dL/dt=\alpha U$$

where α is a constant and $0\leqslant\alpha\leqslant 1$. If $\phi'<0$, then congestion is a problem in that greater numbers of unemployed decrease the rate at which an unemployed individual is absorbed elsewhere in the economy.

The planning objective assumed here is the maximization of the total discounted value of product in the economy, or

$$(6)\qquad \int_0^\infty\{vL+PF(N_x)\}e^{-rt}dt$$

subject to the transfer function, equation (4), where r is the rate of discount. Lapan formally proves that in the finite horizon case an optimal path for unemployment will deviate from the market solution whenever $\phi'\leqslant 0$. Some unemployment will be forestalled completely and the remainder phased over time to reduce congestion problems.

Lapan failed to appreciate, however, the crucial difference in the infinite horizon case between an economy in which $\phi'=0$ and one in which $\phi'<0$. Intuitively it should be clear that if the transfer rate is independent of the size of the transfer ($\phi'=0$), nothing is saved by slow adjustment and therefore full adjustment is optimal instantaneously.

The problem can be conveniently posed as a calculus of variations problem in which equation (6) is maximized subject to the transfer constraint $dL/dt=\alpha U$. Recalling that

$$(7)\quad N_x=L_T-L-U=L_T-L-\frac{1}{\alpha}\left(\frac{dL}{dt}\right)$$

where $L_T\equiv L+L_x$, the total labor in the economy, the Euler condition for an extremal becomes

$$(8)\quad v-\left[1+\frac{r}{\alpha}\right]PF'_x=\frac{P}{\alpha}F''_x\left[\frac{dL}{dt}+\frac{1}{\alpha}\frac{d^2L}{dt^2}\right]$$

A solution to this second-order differential equation is

$$(9)\qquad L(t)=L(0)+U(0)[1-e^{-\alpha t}]$$

The optimal employment at time t outside industry x is the sum of initial employment in that sector and the cumulative flow of unemployed workers from x. The latter is simply the integral from time 0 to time t of the instantaneous rate of transfer of unemployed workers (α) times the exponentially declining (at rate α) stock of unemployment remaining from an initial pulse of unemployment $U(0)$.

The magnitude of the optimal one-time adjustment in employment in x and the corresponding initial pulse in unemployment is determined by the relationship:

$$(10)\qquad PF'_x=\alpha\frac{(v-PF'_x)}{r}$$

The layoff of workers in x instantaneously following a decline in the price of x will optimally cease when the value of produce foregone by laying off an additional worker is equal to the probability the unemployed worker will obtain employment in other sectors of the economy (α) times the capitalized productivity gain of such a reallocation. The marginal products of labor in the two sectors will not be equal in the long run since the model incorporates real transfer costs, namely the opportunity cost of unemployed workers.

If $\phi'<0$, the infinite horizon solution is as discussed by Lapan to spread the release of workers over time to reduce congestion costs. Wage subsidies to industry x should decline monotonically over time. These findings should not be surprising to anyone familiar with the optimal investment literature. In such models, profitable investments should be phased over time only if adjustment costs increase at an increasing rate with the volume of investment. The curvature of the labor transfer function thus becomes an important policy parameter.

B. *The Cyclical Model*

The essence of the congestion model is that the release of excess workers should be phased over time so that unemployment is reallocated from periods of low-transfer rates to periods of high-transfer rates. Not surprisingly, similar reasoning suggests that it may be productive to time the release of excess workers for periods of cyclically high demand if transfer rates are high in such periods. This motivation exists even if the transfer rate with respect to industry unemployment is independent of the level of industry unemployment, i.e., $\phi' = 0$. This assertion will be formally proved in this section.

Perhaps the simplest way to capture formally the notion of cyclical variation in the transfer parameter is to assume that

$$\frac{dL}{dt} = \alpha U \tag{11}$$

where α is a constant. Also assume that following a brief interval of time, Δt, the transfer parameter is expected to rise from α_0 to α_1 and remain at α_1 forever. Such a model is obviously not strictly cyclical, although the results should approach those of a more complete cyclical model as the discount rate rises and/or the period of the cycle lengthens.

How will this time shift in the transfer parameter affect the optimal decision rule in the initial interval? After the period Δt, the optimizing planner will choose the time path for labor transfers developed above. Workers in excess of the number needed to bring the marginal costs and marginal returns of further transfers into equality, equation (10), will be instantly released and then absorbed at the rate α_1 into other sectors of the economy. The optimizing decision then reduces to the choice of the unemployment level during Δt, and therefore to the labor force level at the beginning of the α_1 period. Presumably more workers will be carried over (the initial period unemployment will be less) than would be the case were the transfer parameter in the initial period α_1. One would expect, and indeed it is the case, that unemployment will be less than if the transfer parameter were α_0 throughout since the value of waiting is now positive. Transfers postponed will require less unemployment and foregone product for a given quantity of transfer.

The problem is to find the level of initial period unemployment (U) which will achieve the maximum level of the discounted total product (V) of the workers initially employed in industry x. For convenience, define the discounted value of product to be that at the end of the initial period. Then,

$$\begin{aligned} V \cong\; & (1+r\Delta t)\left[PF(L_x(0)-U)\Delta t\right] \\ & + \int_0^\infty PF(L_x(0)-U_0(\alpha_1))e^{-rt}dt \\ & + v\int_0^\infty \alpha_0 U\Delta t\, dt \\ & + v\int_0^\infty \left[U_0(\alpha_1)-\alpha_0 U\Delta t\right](1-e^{-\alpha t})e^{-rt}dt \end{aligned} \tag{12}$$

or the sum of the value of product in the x industry in the initial and second periods, and the increment to the value of output in the rest of the economy. The term $U_0(\alpha_1)$ represents the volume of workers who would be released immediately were α_1 the transfer rate value throughout. (See equation (10).)

Setting the derivative of V with respect to U equal to zero generates the necessary condition for a maximum, namely

$$-(1+r\Delta t)PF' + \frac{v\alpha_0}{\alpha_1 + r} = 0 \tag{13}$$

Taking the limit as $\Delta t \to 0$ and rearranging terms yields the decision rule for unemployment in the initial period,

$$PF' = \alpha_0\left(v - \frac{\alpha_1}{\alpha_0}PF'\right)\Big/ r \tag{14}$$

A comparison of this rule when $\alpha_1 > \alpha_0$ with that which would hold if $\alpha_1 = \alpha_0$ demonstrates that less workers should be released than would be the case were α_0 maintained throughout. The optimizing government will respond to the higher transfer-rate potential in cyclical boom periods, assuming of course that α does rise in such periods, by postponing the release of excess workers until such periods, at least to some extent. Whether transfer rates do vary procyclically is, of course, an empirical question.

The preceding analysis could be undertaken in the alternative situation in which the economy is predicted to move from a high level of labor demand to a low level. In such a case, the government planner may wish to encourage immediate unemployment through some employment tax policy. The extent of such a policy will depend in part on the nature of rehire possibilities and will not be pursued at this time.

II. Determinants of the Rate of Transfer: An Empirical Analysis

The theoretical analysis of the preceding section suggests that optimal-unemployment-motivated transfers will depend on the shape of the labor transfer function. Are unemployed workers absorbed at a slower rate if the volume of industry unemployment is higher? If the economy-wide unemployment rate is higher? The former process would argue for smoothing the release of "excess" workers over time. The latter would indicate that the release of such workers may be optimally postponed until cyclical peak periods. In this section, I report on estimates of the parameters of the transfer function obtained from aggregate and disaggregate manufacturing industry data monthly over the period January 1968 to January 1976.

The essence of the approach is to obtain estimates of the factors which affect the rate of retention in unemployment status of unemployed job losers who are not recalled to their jobs. The BLS publishes monthly data on job turnovers by industry. In particular, the BLS reports on the extent of layoffs, defined as "employer initiated terminations without prejudice to the worker" and reports on total accessions and new hires, the latter defined as the addition of workers not previously employed in the firm. The difference between these two accession figures yields a measure of rehires or recalls. The BLS also collects, but does not publish, monthly unemployment stock measures by cause for aggregate manufacturing and for selected 2-digit manufacturing industries. From these it is possible to obtain monthly measures of unemployment among job losers.

The stock and flow measures together provide a method of obtaining estimates of retention rates in unemployment among job losers if they are not recalled. One minus the fractional retention rate can be interpreted as the transfer rate of job losers in the industry to other firms. If the retention rate is a constant, it can be identified by the coefficient γ_1 of the following difference equation,

$$(15)\quad UR = \gamma_0 + \gamma_1 UR(-1) + L - R + \varepsilon$$

where $UR \equiv$ unemployment rate among job losers
$L \equiv$ layoff rate
$R \equiv$ recall rate
$\varepsilon \equiv$ random error term

Any reduction in the number of unemployed job losers not accounted for by net inflows (layoffs minus recalls) to unemployment are assumed to be absorbed elsewhere in the economy.[7] The rate of transfer among unemployed job losers is then $(1-\gamma_1)$.

The discussion of the previous section indicates the importance of determining whether the retention rate is constant over varying economic conditions. The model estimated below has the form

$$(16)\quad UR = \gamma_0 + \gamma_1 UR(-1) \cdot RET + L - R + \varepsilon$$

---

[7] One possible transfer (into the household) has different potential policy implications than other transfers. In the empirical work below it is, unfortunately, impossible to distinguish between those who leave unemployment because they find other jobs and those who disappear from unemployment because they leave the labor force.

where $RET \equiv$ the retention rate function. The retention rate is potentially a function of the unemployment rate in the industry (the congestion effect), the unemployment rate in the rest of the economy, and perhaps the probability of future recall. Feldstein (1975, 1976) in particular has discussed the importance of recall in affecting job search intensity. See also my paper. A linear retention rate model then would be

$$RET = \beta_0 + \beta_1 UR(-1) + \beta_2 URE + \beta_3 RPROB \tag{17}$$

where $URE \equiv$ unemployment rate in the remainder of the economy and $RPROB \equiv$ recall probability. For the congestion argument for government intervention to hold, presumably $\beta_1 > 0$; for the cyclical smoothing, argument, $\beta_2 > 0$.

Since the BLS turnover data suffer from a well-known problem of scale,[8] the assumption that the coefficients of layoffs and rehires should be one and minus one, respectively, in equation (16) seems unwarranted. This problem precludes the simple calculation of *RET* from equation (16) which would then be the dependent variable in equation (17). Instead the *RET* expression in equation (17) was substituted into equation (16) and the elaborated version of (17) was then estimated.

The estimation technique employed is ordinary least squares. Experimentation on the aggregated industry data with Cochrane-Orcutt techniques of adjusting for first-order autocorrelation provided little evidence of first-order autocorrelation and only minor variations in coefficient estimates. The Durbin-Watson statistics reported below, although biased toward acceptance of the null hypothesis of no autocorrelation, are apparently useful approximations.[9]

The results of the estimation of the job-loser unemployment rate equations for manufacturing and durable and nondurable manufacturing are reported in Table 1. In the first row of Table 1, the simple autoregressive model of the job-loser unemployment rate is reported for manufacturing. The coefficient on the lagged value of the unemployment rate indicates that 89 percent of such unemployed remain unemployed over a one-month interval if they are not recalled. The remaining 11 percent are absorbed elsewhere in the economy. The estimates for durable goods industries and nondurable goods industries, reported in the fourth and seventh rows, respectively, are virtually identical.

The congestion model assumes that the retention rate rises with higher levels of industry unemployment. Inclusion of a squared lagged unemployment rate term provides no support for this proposition. The regression model with the squared term included is reported for manufacturing in the second row of Table 1, for durable manufacturing in the fifth row, and for nondurables in the eighth row. Only for durable goods is the squared term coefficient positive and in no case is it significant.

The full model is estimated in the third row for manufacturing. Both the economy-wide unemployment rate and the probability of recall increase retention in unemployment status with statistically significant estimates. The probability of recall is measured as the ratio of unemployed job losers on layoff to all unemployed job losers. Job losers who report themselves on layoff are presumed to have higher recall likelihood than "other job losers," for example, those resulting from permanent plant closing, etc.

The unemployment and recall probability effects are absolutely large as well as statistically significant. A rise in the economy-wide unemployment from .04 to .08 will increase the monthly retention rate among unemployed job losers by 7.7 percentage points. Given an average transfer rate of 10 percent, this is obviously a variation of considerable consequence which may justify wage subsidies to contracting industries during cyclical contradictions.

The effect of recall probability on the retention rate is also large. A one percentage

---

[8] In particular, the summation of net employment turnovers when added to initial stocks of the employment systematically underestimates directly measured employment stocks.

[9] See Jack Johnston (pp. 311-12).

TABLE 1—UNEMPLOYMENT REGRESSIONS, AGGREGATE U.S. MANUFACTURING INDUSTRIES
(Monthly 1968–76)

Constant	$UR(-1)$[a]	$[UR(-1)]^2$	$UR(-1)$ $\cdot URE$[b]	$UR(-1)$ $\cdot RPROB$[c]	Layoff[d]	Rehire[e]	R^2	$D.W.$[f]
All Manufacturing								
−0.0055	0.888				0.947	−0.254	0.978	2.01
(3.48)	(37.86)				(14.14)	(1.64)		
−0.0060	0.908	−0.185			0.946	−0.243	0.978	2.01
(2.76)	(14.31)	(0.34)			(14.02)	(1.53)		
−0.0056	0.721	−2.425	1.924	0.499	0.779	−0.215	0.980	1.95
(2.61)	(7.26)	(2.22)	(2.12)	(2.12)	(7.29)	(1.37)		
Durable Manufacturing								
−0.0044	0.890				0.948	−0.297	0.978	2.06
(2.79)	(34.09)				(13.76)	(1.75)		
−0.0034	0.848	0.0367			0.953	−0.314	0.978	2.08
(1.62)	(12.93)	(0.69)			(13.72)	(1.82)		
−0.0043	0.675	−2.260	2.954	0.360	0.840	−0.207	0.980	2.00
(2.09)	(7.96)	(2.10)	(2.91)	(1.77)	(8.12)	(1.21)		
Nondurable Manufacturing								
−0.0076	0.880				0.933	−0.123	0.957	2.24
(3.20)	(30.85)				(10.38)	(0.67)		
−0.0100	0.967	−0.881			0.940	−0.073	0.957	2.26
(2.97)	(10.70)	(1.02)			(10.43)	(0.38)		
−0.0081	0.716	−3.12	1.968	0.587	0.769	−0.095	0.960	2.19
(2.24)	(4.62)	(2.41)	(1.49)	(2.42)	(6.80)	(0.50)		

Sources: BLS Employment and Earnings data tape (see BLS, 1972), various issues of *Employment and Earnings*, and unpublished BLS data.

Note: Dependent variable is the experienced worker unemployment rate (unemployment end of month divided by the sum of unemployment end of month and employment midmonth) among job losers. Absolute values of the *t*-ratios are in parentheses. The data cover ninety-seven months, January 1968 to January 1976.

[a] The dependent variable, lagged once.

[b] The economy-wide unemployment rate excluding employment and unemployment in the industry being modelled.

[c] A proxy for recall probability, namely the ratio of job losers on layoff to all unemployed job losers.

[d] The ratio of BLS layoffs to total employment.

[e] The ratio of recalls (BLS accessions minus new hires) to total employment.

[f] The Durbin-Watson is biased toward acceptance of the null hypothesis in the presence of lagged endogenous variables although it may provide a reasonably good indicator even in this case. See fn. 9.

point increase in the recall probability increases the retention rate by one-half point. The expectation of recall apparently provides a powerful incentive for workers to inventory themselves in unemployment.

The coefficient on the squared $UR(-1)$ coefficient is negative and statistically significant, however, suggesting that the retention rate drops as the industry's stock of unemployed job losers rise. This finding is somewhat puzzling, given the intuitive plausibility of the congestion model. The powerful effect of recall likelihood may partially explain this finding. As the stock of unemployed expands, the likelihood of timely recall no doubt drops, even holding constant the permanence of the job loss, thus inducing workers to accept jobs elsewhere that they otherwise might forego. Barron finds some evidence that unemployed workers reduce their wage expectations when unemployment levels are high.

The coefficient estimates on the layoff and rehire variables are of the expected signs but implausibly low for rehires. This may reflect simultaneity bias in the estimate since rehires will be low in periods when the stock of unemployed workers is low. The stock of unemployed workers constrains the extent of rehire. Reestimation of the full model constraining the coefficients on layoffs and rehires to be 0.8 and −0.8, respectively, did

TABLE 2—UNEMPLOYMENT REGRESSIONS, DISAGGREGATED MANUFACTURING INDUSTRIES
(Monthly 1968–76)

Industry	Constant	$UR(-1)$	$[UR(-1)]^2$	$UR(-1) \cdot URE$	$UR(-1) \cdot RPROB$	Layoff	Rehire	R^2	$D.W.$
				Durable Goods Industries					
Primary Metals	−.001325	.5439	−1.238	3.639	.05909	.8827	.05598	.9522	2.36
	(−.52)	(3.14)	(−1.06)	(3.23)	(.33)	(9.30)	(.26)		
Fabricated Metals	−.004849	.2184	−5.275	9.505	.4002	.7140	.6062	.9210	2.13
	(−1.25)	(1.23)	(−3.38)	(4.66)	(1.86)	(5.42)	(1.99)		
Machinery	−.00364	.7499	−3.143	3.291	.1480	.7422	.1822	.9205	2.35
	(−1.16)	(4.79)	(−1.53)	(2.31)	(.65)	(3.50)	(0.48)		
Electrical Equipment	−.0008121	.4145	−1.362	5.246	.1718	.9587	.08086	.963	2.19
	(−.26)	(2.82)	(−1.10)	(2.63)	(.86)	(7.08)	(.28)		
Autos	.01263	−.8213	−1.787	9.970	1.508	.6543	−.3340	.884	2.03
	(2.44)	(−2.99)	(−2.78)	(5.29)	(4.97)	(7.34)	(−1.87)		
Other Transportation Equipment	−.01887	.7743	−2.067	1.209	.2066	1.041	.8405	.8739	2.58
	(−3.18)	(3.62)	(−1.35)	(.94)	(.94)	(5.62)	(2.61)		
Other Durable Goods	−.005059	.6067	−3.019	2.619	.7671	.7115	−.007464	.9489	1.76
	(−1.57)	(4.92)	(−2.64)	(2.08)	(2.73)	(5.74)	(−.002)		
				Nondurable Goods Industries					
Food and Kindred Products	.01056	.7919	−3.844	4.018	.3054	.2662	−.8100	.8471	2.24
	(1.15)	(3.65)	(−1.68)	(2.52)	(1.14)	(1.90)	(−3.31)		
Textiles	−.01321	.4092	−2.557	5.552	.07007	1.438	1.089	.9253	2.16
	(−2.45)	(1.95)	(−2.64)	(2.86)	(.33)	(8.95)	(2.66)		
Apparel	.004212	−.03734	−1.734	6.890	.8000	.5772	−.01881	.8715	1.81
	(.62)	(−.15)	(−1.67)	(3.72)	(3.58)	(4.33)	(−.08)		

Notes: Dependent variable is the experienced worker unemployment rate among job losers. The absolute value of the *t*-ratios are in parentheses. For data sources and variable definitions, see Table 1.

not, however, substantially alter the signs or statistical significance of the retention rate variables.

The estimates of the full model for durable and nondurable manufacturing yielded results comparable to those for all manufacturing. These are reported in the sixth and ninth rows, respectively, of Table 1. The robustness of the model under disaggregation, however, is most apparent in Table 2, in which estimates of the full model are reported for selected 2-digit manufacturing industries. For each of these industries the signs and significance of the retention rate variables are the same as in the aggregate estimates. As might be expected, the unemployment rate in the rest of the economy has generally a larger magnitude of effect in the more narrowly defined industries. The quadratic term in lagged unemployment remains uniformly negative, although not always significantly so.

III. Conclusion

This paper has investigated theoretically and empirically unemployment-induced labor transfers among firms. With wage rates relatively inflexible, reallocation among firms is largely induced by unemployment. In Section II, I demonstrated formally that at least two reasons may exist for temporary government intervention in the labor market to alter the time path of firm layoffs in industries with declining sales. If higher layoff rates result in congestion in local labor markets, so that the rate of transfer of workers to other firms is reduced, the government may optimally choose to spread the release of excess workers over time by temporarily subsidizing continued employment in the adversely affected industry. If high levels of unemployment elsewhere in the economy reduce the transfer rate of the industry's unemployed workers to other firms, the government may also optimally intervene to postpone the release of workers until periods of higher general labor demand.[10]

The rationale for intervention in each case rests on the empirical proposition that industry and economy-wide unemployment affects the transfer rate to other sectors. Using monthly unemployment and turnover data by industry over the period 1968–77, the characteristics of the transfer function are estimated. The results indicate no "congestion" effects of the industry's own unemployment level.[11] If anything, the rate of transfer rises with industry unemployment, perhaps as the result of the reduced likelihood of rapid recall to the previous firm. A more direct measure of recall probability suggests that the recall likelihood effect is indeed a powerful factor in the determination of the transfer rate of unemployed workers. The economy-wide unemployment rate did reduce the rate of transfer of unemployed job losers elsewhere in the economy, which would support the notion of cyclical shifting of layoffs to periods of high demand.

---

[10] Both of these conjectures ignore any general equilibrium effects of a more or less rapid rate of reallocation. As an anonymous referee has pointed out, the rate of recovery after a period of low aggregate demand may be influenced by the rate of adjustment to a new equilibrium and by the extent of unemployment during the transition.

[11] As an anonymous referee notes, the unemployment compensation system already taxes permanent layoffs (which would have an effect equivalent to subsidizing employment) to the extent the system has at least partial experience rating.

REFERENCES

A. Alchian, "Information Costs, Pricing, and Resource Unemployment," in Edmund S. Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970, 27–52.

C. Azariadis, "Implicit Contracts and Underemployment Equilibria," *J. Polit. Econ.*, Dec. 1975, *83*, 1183–202.

M. N. Baily, (1974a) "Wages and Employment Under Uncertain Demand," *Rev. Econ. Stud.*, Jan. 1974, *41*, 37–50.

______, (1974b) "Unemployment and Unemployment Insurance," mimeo., Yale Univ. 1974.

______, "Unemployment Insurance as Insurance for Workers," mimeo., Yale Univ. 1975.

W. F. Barnes, "Job Search Models, the Duration of Unemployment, and the Asking Wage: Some Empirical Evidence," *J. Hum. Resources*, Spring 1975, *10*, 230–40.

J. M. Barron, "Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence," *Amer. Econ. Rev.*, Dec. 1975, *65*, 934–42.

F. Brechling, "The Incentive Effects of the U.S. Unemployment Insurance Tax," in Ronald Ehrenberg, ed., *Research in Labor Economics*, Vol. 1, Greenwich 1977, 41–102.

______, *Investment and Employment Decisions*, Manchester 1975.

R. G. Ehrenberg and R. L. Oaxaca, "Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain," *Amer. Econ. Rev.*, Dec. 1976, *66*, 754–66.

M. S. Feldstein, "Temporary Layoffs in the Theory of Unemployment," *J. Polit. Econ.*, Oct. 1976, *84*, 937–58.

______, "The Importance of Temporary Layoffs," *Brookings Papers*, Washington 1975, *3*, 725–44.

D. F. Gordon, "A Neo-Classical Theory of Keynesian Unemployment," *Econ. Inquiry*, Dec. 1974, *12*, 431–59.

H. I. Grossman, "The Nature of Optimal Labor Contracts," mimeo., Brown Univ. 1975.

J. Harris and M. Todaro, "Migration, Unemployment, and Development: A Two-Sector Analysis," *Amer. Econ. Rev.*, Mar. 1970, *60*, 126–42.

A. Holen and S. A. Horowitz, "The Effect of Unemployment Insurance and Eligibility Enforcement on Unemployment," *J. Law Econ.*, Oct. 1974, *17*, 403–32.

H. G. Johnson, "Optimal Trade Intervention in the Presence of Domestic Distortions," in Richard E. Caves et al., eds., *Trade, Growth, and the Balance of Payments*, Amsterdam 1965.

Jack Johnston, *Econometric Methods*, 2d ed., New York 1972.

H. B. Kaitz, "Analyzing the Length of Spells of Unemployment," *Mon. Labor Rev.*, Nov. 1970, *93*, 11–20.

H. Kasper, "The Asking Price of Labor and the Duration of Unemployment," *Rev. Econ. Statist.*, May 1967, *49*, 165–73.

H. E. Lapan, "International Trade, Factor Market Distortions, and the Optimal Dynamic Subsidy," *Amer. Econ. Rev.*, June 1976, *66*, 335–46.

S. Magee, "Factor Market Distortions, Production, and Trade: A Survey," *Oxford Econ. Papers*, Mar. 1973, *25*, 1–43.

S. McCafferty, "Optimal Wage and Layoff Policy in Stochastic Equilibrium," mimeo., 1978.

D. O. Parsons, "Experienced Worker Unemployment Behavior," mimeo., 1976.

G. L. Perry, "Unemployment Flows in the U.S. Labor Market," *Brookings Papers*, Washington 1972, *2*, 245–92.

Edmund S. Phelps, *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

R. Schmidt, "The Theory of Search and the Duration of Unemployment," work. paper, Univ. Rochester 1973.

A. B. Treadway, "On Rational Entrepreneurial Behavior and the Demand for Investment," *Rev. Econ. Stud.*, Apr. 1969, *36*, 227–39.

U.S. Bureau of Labor Statistics, *Employment and Earnings, United States, 1909–72*, Washington 1972.

______, *Employment and Earnings*, various issues.

______, "Unemployment Tables—Monthly, Experienced Unemployed by Industry, Age, Sex, and Color," unpublished monthly data, 1968–1976.

Fiscal Policies, Inflation, and Capital Formation

By MARTIN FELDSTEIN*

The large unprecedented government deficits in recent years have stimulated speculation about their adverse effects on inflation and private capital formation. While it is clear that deficits may have no adverse effect in an economy with sufficient unemployed resources, the effects of a deficit when there is full employment are less clear. Is a persistent increase in the government deficit necessarily inflationary? Does it necessarily reduce private capital formation? Is it possible to avoid both adverse effects? A primary purpose of this paper is to answer these questions in the context of a fully employed and growing economy.

A closely related issue is the relation between private saving and capital formation when money and other government liabilities are alternatives to real capital in individual portfolios. John Maynard Keynes, Roy Harrod, and James Tobin have all emphasized the possibility of excess saving when individuals will not hold capital unless its yield exceeds some minimum required return. When the return on capital is too low, an increase in saving only reduces aggregate demand. If prices are flexible downward, this causes deflation until the increased value of real balances causes a sufficient reduction in saving; if prices cannot fall, the excess saving results in unemployment.

Three ways of averting such "excess saving" have been emphasized in both theory and practice. The thrust of the Keynesian prescription was to increase the government deficit to provide demand for the resources that would not otherwise be used for either consumption or investment. In this way, aggregate demand would be maintained by substituting public consumption for private consumption. A second alternative prescription was to reduce the private saving rate. Early Keynesians like Seymour Harris saw the new Social Security program as an effective way to reduce aggregate saving. The third type of policy, developed by Tobin, relies on increasing the rate of inflation and making money less attractive relative to real capital. In Tobin's analysis, the resulting increase in capital intensity offsets the higher saving rate and therefore maintains aggregate demand.

This paper will examine ways of increasing capital intensity without raising the rate of inflation. The analysis will also show why, contrary to Tobin's conclusion, a higher rate of inflation may not succeed in increasing investors' willingness to hold real capital.

An important feature of the analysis in this paper is a monetary growth model that distinguishes between money and interest-bearing government bonds. With this distinction, we can compare government deficits financed by borrowing with deficits financed by creating money. It is possible also to examine the effect of changes in the interest rate on government debt while maintaining the fact that money is not interest bearing. The two types of government liabilities also permit analyzing the distinction between the traditional liquidity preference and a demand for government bonds that I shall call safety preference. In practice, this safety preference may be much more important than the traditional liquidity preference.

The first section of the paper develops the three-asset monetary growth model that will be used in the remaining analysis. Section II then considers the effects of changes in the government deficit. The effects of increased saving on aggregate demand and capital intensity are developed in Section III.

---

*President, National Bureau of Economic Research, and professor of economics, Harvard University. This paper is part of the NBER Program of Research on Capital Formation; a previous version was circulated as NBER working paper no. 275 (Aug. 1978). I am grateful for financial support from the National Science Foundation and the NBER. The paper presents my views and not those of the NBER.

I. A Three-Asset Model of Monetary Growth

The model developed here differs from the traditional monetary growth (see, for example, Tobin, and David Levhari and Don Patinkin) in two important ways. First, instead of the usual assumption that all taxes are lump sum levies, the current analysis recognizes taxes on capital income that lower the net rate of return.[1] Second, the government deficit is financed not only by increasing the money supply but also by issuing interest-bearing government debt.[2] Throughout the analysis we maintain the simplifying assumption that the savings rate out of real disposable income is fixed; the tax on capital income affects the allocation of saving but not the saving rate itself. Because the analysis in the sections that follow will focus on comparative steady-state dynamics, only these steady-state properties will be discussed here.[3] Note that this assumption of steady-state dynamics implies that all growth rates are constant and therefore that all quantities are correctly anticipated.

The economy is characterized by an exogenously growing population

$$N=N_0e^{nt} \tag{1}$$

The labor force is a constant fraction of the population and technical progress is subsumed into population growth. Production can be described by an aggregate production function with constant returns to scale. The relation between per capita output (y) and the per capita capital stock (k) is

$$y=f(k) \tag{2}$$

with $f'>0$ and $f''<0$. For simplicity, output is measured net of depreciation and depreciation is not explicitly included in the analysis.

A. *The Government Budget Constraint*

Government spending (G) plus the payment of interest on the government debt must be financed by either tax receipts, money creation, or borrowing. Total real tax ceipts (T) are the sum of a lump sum tax (T_0) and the revenue that results from taxing the income from real capital at rate τ.[4] The money created by the government (M) is the only money in the economy and does not bear interest. The time rate of change of the stock of nominal money is DM; the real value of the extra money created in this way is DM/p. Government bonds bear interest at rate i; the nominal market value of these bonds is B and the real value of new borrowing is DB/p.[5]

The government's budget constraint may be written

$$G+\frac{iB}{p}=T+\frac{DM}{p}+\frac{DB}{p} \tag{3}$$

---

[1] My 1976 paper and my 1978 paper with Jerry Green and Eytan Sheshinski show the importance of recognizing capital income taxes in analyzing the effects of inflation in a monetary growth model. Corporate and personal taxes were distinguished there but will not be in this paper.

[2] Green and Sheshinski examine an economy with both bonds and money but assume that such bonds are perfect substitutes for private capital in investors' portfolios. Their analysis generally focuses on quite different issues. Tobin and Willem Buiter recently developed a three-asset model in which government bonds and money are imperfect substitutes for each other as well as for real capital. I received a copy of their unpublished paper only after this paper was submitted for publication. Because the tax structure that they assume is very different from the taxes described herein, their conclusions are frequently different from my own. Benjamin Friedman (1978) also recently developed a three-asset model but used it to analyze quite different questions (the short-run effects of monetary and fiscal policies in an economy with unemployment and fixed prices) from those that are the focus of my own research.

[3] Section III will, however, consider the possibility of disequilibrium behavior associated with excess saving.

[4] The tax rate τ is best thought of as a corporate tax rate. The personal tax on real capital income and on the interest on government debt is not specially recognized.

[5] These bonds may be thought of as Treasury bills although their maturity is irrelevant for steady-state analysis as long as that maturity is finite. I ignore changes in the market value that would temporarily result from changes in the interest rate if the maturity were not very short.

Alternatively, it will be convenient to denote the real government deficit by Δ and write

$$\Delta=\frac{DM}{p}+\frac{DB}{p} \tag{4}$$

In the steady state, the ratio of real money per unit of real capital (M/pK) must remain constant. This implies that the rate of growth of M is equal to the rate of growth of pK, or with $Dp/p=\pi$,[6]

$$\frac{DM}{M}=\pi+n \tag{5}$$

Similarly, the steady-state rate of growth of nominal government bonds equals the inflation rate plus the real growth rate of the economy:

$$\frac{DB}{B}=\pi+n \tag{6}$$

Substituting these expressions into (4) and dividing by the population gives the steady-state per capita deficit:

$$\frac{\Delta}{N}=(\pi+n)\frac{M}{pN}+(\pi+n)\frac{B}{pN} \tag{7}$$

With lowercase letters representing real per capita values, (7) can be rewritten

$$\delta=(\pi+n)(m+b) \tag{8}$$

The real per capita deficit equals the product of the economy's nominal growth rate and the real per capita government liabilities.

B. *Portfolio Behavior*[7]

The real value of household assets is the sum of the real values of government liabilities and the capital stock:[8]

$$a=b+m+k \tag{9}$$

I shall simplify the description of the households' portfolio behavior by assuming that the equilibrium ratio of real bonds to capital depends on the difference between the net real yield on capital (r) and the real yield on government bonds ($i-\pi$):[9]

$$\frac{b}{k}=\beta[r+\pi-i] \qquad \beta'<0 \tag{10}$$

Because the depreciation method used in the United States and in most other countries is based on the original costs of plant and equipment, the tax liability per unit of capital increases with the rate of inflation.[10] I shall therefore write the net rate of return as

$$r=f'-\tau(f'+\lambda\pi) \tag{11}$$

where the parameter λ indicates the extent to which a higher inflation rate increases the tax liability.[11] Substituting into (10) yields the equilibrium bond portfolio condition:

$$\frac{b}{k}=\beta[(1-\tau)f'+(1-\tau\lambda)\pi-i] \qquad \beta'<0 \tag{12}$$

---

[6] This uses the fact that in the steady state $k=K/N$ is constant, implying $DK/K=n$.

[7] The description of portfolio behavior and saving in this model assumes that households as a whole regard government bonds as net wealth, implicitly ignoring the corresponding tax liabilities that they and future generations must bear in order to pay the interest and principle on these bonds. The alternative view, based on the so-called "Ricardian equivalence theorem," depends on extremely strong and improbable assumptions. To the extent that households do partially reduce their perceived value of government bonds because of future tax liabilities, the variable B (or b) in household portfolios might be scaled down to some fraction of B. This need not change any of the qualitative conclusions of the current analysis.

[8] The private bonds and equities that represent the ownership claims to the capital stock are not explicitly recognized. The tax rate τ can be regarded as the effective tax rate corresponding to the steady-state mix of debt and equity finance. See my 1979 paper with Green and Sheshinski.

[9] Recall that we are ignoring the *personal* tax on investment income.

[10] See my 1978 paper with Green and Sheshinski, especially the appendix by Alan Auerbach, and my 1977 paper with Lawrence Summers.

[11] Accelerated depreciation affects τ and λ but λ exceeds zero even if, at the equilibrium value of π, tax depreciation exceeds economic depreciation.

With safe short-term interest-bearing government debt available, individuals should hold money only for transaction purposes. As William Baumol and others have shown, this demand for money varies positively with the level of income and inversely with the rate of interest:[12]

$$\frac{m}{y}=L(i),\ L'<0 \tag{13}$$

An important feature of an economy with money or other government liabilities is the possibility that individuals will be unwilling to hold capital unless its yield is above some minimum level. In the traditional two-asset Keynesian model, this is represented as a liquidity trap, that is, as an infinitely elastic demand for money at some low rate of interest. A more realistic description is possible with the current three-asset model. When the real net yield on capital becomes very low relative to the real yield on government bonds, investors will want to hold government bonds instead of capital; in the notation of equation (12), the absolute value of β' becomes infinite when the real differential becomes very small.[13] The reason that investors prefer government bonds in this situation is that the pretax profitability of private capital is uncertain. The bond-demand behavior will therefore be referred to as a "safety preference" relation to distinguish it from the liquidity preference relation that governs the demand for money.[14]

C. *The Supply and Demand for Savings*

The supply of savings (S) is proportional to the households' real disposable income (H):

$$S=\sigma\cdot H \tag{14}$$

The saving propensity will be assumed to be constant. Disposable income is equal to national income (Y) minus both the government's tax receipts (T) and the fall in the real value of the population's money and government bonds ($\pi M/p$ and $\pi B/p$).[15] Saving is therefore

$$S=\sigma\left(Y-T-\frac{\pi M}{p}-\frac{\pi B}{p}+\frac{iB}{p}\right) \tag{15}$$

or, using equation (3),

$$S=\sigma\left(Y-G+\frac{DM}{p}+\frac{DB}{p}-\frac{\pi M}{p}-\frac{\pi B}{p}\right) \tag{16}$$

Since $DM/p-\pi M/p=nM/p$ (with a similar equivalence for bonds),

$$S=\sigma\left(Y-G+\frac{nM}{p}+\frac{nB}{p}\right) \tag{17}$$

In the steady state, government spending must bear a stable relation to national income. The analysis that follows assumes that a fraction γ of national income is devoted to government spending exclusive of interest on the government debt. This implies that any increase in interest on the government debt causes a corresponding increase in the government deficit or in lump sum taxes.

All saving must be absorbed in either real capital accumulation or additional real money and bonds:

$$S=DK+D(M/p)+D(B/p) \tag{18}$$

---

[12] This simplifies by assuming that individuals regard the interest-bearing government debt rather than real capital as an alternative to transaction balances. Transaction balances are also assumed to depend on income rather than wealth when in reality both are important.

[13] This unwillingness to own real capital may also increase the demand for money, but that effect is likely to be small relative to the increased demand for bonds.

[14] The private securities are generally as marketable as government bonds and, to the extent that they have the same maturity structure, their price will be as sensitive to interest rate fluctuations. Their liquidity is therefore similar even though the safety and predictability of the yields differ substantially.

[15] Recall that this analysis assumes that households as a whole regard government bonds as net wealth.

The constant ratio of capital to labor in steady-state growth implies $DK=nK$. Similarly, the constancy of $m=M/pN$ and $b=B/pN$ implies that $D(M/p)=nM/p$ and $D(B/p)=nB/p$. Thus

$$S=nK+\frac{nM}{p}+\frac{nB}{p} \tag{19}$$

Combining equations (17) and (19), writing γY for G, and dividing by N yields the per capita growth equilibrium condition:

$$\sigma[y(1-\gamma)+nm+nb]=nk+nm+nb \tag{20}$$

This completes the presentation of the model that will be analyzed in the remainder of this paper. Although the model contains important features that were lacking in earlier monetary growth models, it is still very simple. It would be desirable to investigate a richer class of models that includes personal taxes on portfolio income, that distinguishes households from institutional investors, that recognizes substitutes for business capital other than bonds and money, and that separates corporate bonds and equity. The reader should bear in mind the strong simplifying assumptions that have been made, and the fact that the subsequent analysis will compare steady-state growth patterns with no attention to the transitions. Any such short-run or transitional analysis raises difficult problems of specifying how expectations are formed when the economy is not on a steady-state growth path.

II. Deficits, Inflation, and Capital Intensity

The model developed in Section I can now be used to analyze how changes in the government deficit affect the rate of inflation and the capital intensity of the economy. Can the government increase the real steady-state deficit in this fully employed economy without causing either inflation, reduced capital intensity, or both? What policies can be pursued to mitigate the adverse effects of the deficit? What happens when the policy options of the government are restricted?

To answer these questions, it is useful to collect the four equations that describe the steady-state behavior of the economy:

$$\delta=(\pi+n)(m+b) \quad \text{(previously (8))} \tag{21}$$

$$m=L(i)\cdot f(k) \quad \text{(previously (13))} \tag{22}$$

$$b=\beta[(1-\tau)f'(k)+(1-\tau\lambda)\pi-i]\cdot k \quad \text{(previously (12))} \tag{23}$$

$$\sigma[(1-\gamma)f(k)+nm+nb] = n[k+m+b] \quad \text{(previously (20))} \tag{24}$$

where $f(k)$ has been substituted for y in (22) and (24). The policy instruments controlled by the government are the size of the deficit (δ), the share of government spending in national income (γ), the interest rate on government bonds[16] (i), and the tax rates on capital income (τ and λ). For given values of these policy instruments and the exogenous growth rate (n), the four equations determine the values of k, π, m, and b.

It is clear from these four equations that the government can increase its deficit without inducing any changes in inflation or capital intensity if it can vary all of the other policy instruments (γ, i, τ, and λ). In practice, however, the government does not alter the share of government spending in national income (γ) in order to neutralize the effect of a deficit.[17] It is tempting to conclude that, even if γ is held constant, the government can still increase the deficit without changing π or k because it still has three unconstrained instruments. It is easily shown, however, that this is not true; an increased deficit must then be accompanied by a change in either inflation, capital intensity, or both. To see this note that (with γ constant) equation (4) implies that if $dk=0$ it is also true that $d(m+b)=0$. Equation (21) shows that $d(m+b)=0$ and $d\pi=0$ together imply $d\delta=0$. The deficit must be unchanged if both inflation and capital intensity are unchanged.

---

[16] Recall that this would not in general equal the rate of return on private capital. This interest rate could equivalently be regarded as endogenous if the government is assumed to choose the supply of nominal bonds, or the supply of money, or the ratio of these two.

[17] This would in particular require reducing the share of government spending in national income. To see this, note that equation (21) implies $d\delta=(\pi+n)d(m+b)$ if $d\pi=0$. Equation (24) implies $-k\sigma f d\gamma=(1-\pi)nd(m+b)$ since $dk=0$. Combining these two shows $d\gamma/d\delta<0$.

The government can, however, affect the combination of changes in inflation and capital intensity that occurs by its debt-management policy and its tax policy. Because changes in tax policy (in τ and λ) are not a typical government response, most of this section will assume that τ and λ as well as γ are unchanged. My analysis focuses on debt-management policy, that is, on the way that the government adjusts the relative supply of money and bonds or, equivalently, the rate of interest on government debt.

A. *A Deficit Causing Both Increased Inflation and Reduced Capital Intensity*

With the type of debt policy that has been pursued in the United States, an increase in the steady-state deficit is likely to cause both a higher rate of inflation and a reduced capital intensity of production. More specifically, most empirical research indicates that the government has issued a mix of money and debt in such a way that the *real* interest rate on government debt remains approximately unchanged.[18] This section analyzes the consequences of this policy; alternative debt policies are examined in the subsequent section. To analyze the effect of an increased deficit with a constant *real* interest rate on government debt, equations (21)–(24) are totally differentiated subject to the condition $di=d\pi$.

The key to the adverse effect of inflation on capital intensity is seen in the total differential of (23), subject to $di=d\pi$:

$$db=(\beta+k\beta'(1-\tau)f'')dk-k\beta'\tau\lambda d\pi \tag{25}$$

The partial effect of an increase in inflation is to increase the demand for bonds rather than capital because the real yield on bonds is maintained while the real yield on capital falls by $\tau\lambda d\pi$. If this positive effect of inflation on the demand for bonds is large enough to outweigh the negative effect of inflation on the demand for money that is implied by equation (22) with $di=d\pi$:

$$dm=Lf'dk+fL'd\pi \tag{26}$$

the effect of an increased deficit can be shown unambiguously to reduce k. To see this, note first that (24) implies

$$d(m+b)=\frac{\sigma(1-\gamma)f'-n}{(1-\sigma)n}dk \tag{27}$$

Adding (25) and (26) and then using (27) to eliminate $d(m+b)$ yields

$$\frac{(1-\gamma)f'-n}{(1-\sigma)n}dk=(Lf'+\beta+k\beta'(1-\tau)f'')dk+(fL'-k\beta'\tau\lambda)d\pi \tag{28}$$

Similarly, differentiating equation (21) and using (27) yields

$$d\delta=\frac{\sigma(1\ \ \gamma)f'-n}{(1-\sigma)n}dk+(m+b)d\pi \tag{29}$$

Using (28) to eliminate $d\pi$ from (29) yields

$$\left.\frac{dk}{d\delta}\right|_{di=d\pi}=[(1-\sigma)n(fL'-k\beta'\tau\lambda)]/ \tag{30}$$
$$[(\pi+n)(\sigma(1-\gamma)f'-n)(fL'-k\beta'\tau\lambda)$$
$$+(m+b)(\sigma(1-\gamma)f'-n-n(1-\sigma)$$
$$\times(Lf'+\beta+k\beta'(1-\tau)f''))]$$

---

[18] Evidence that the nominal interest rate rises by approximately the rate of inflation was presented by Irving Fisher and has been verified by William Yohe and Denis Karnovsky, my paper with Otto Eckstein, my 1977 paper with Summers, and others. The assumption in Tobin that $di=0$ is clearly inconsistent with experience when i is interpreted as the yield on government debt rather than the yield on money.

With the increased demand for bonds induced by higher inflation greater than the reduced demand for money, $fL'-k\beta'\tau\lambda > 0$, and the numerator is positive. Note that stability of the simpler nonmonetary economy requires $\sigma(1-\gamma)f'-n < 0$; with this assumption, the denominator is negative. Under these quite plausible conditions, a higher deficit reduces capital formation. Since (28) implies that dk and $d\pi$ are of opposite signs, the higher deficit also increases inflation.

B. *A Deficit without Inflation*

The bleak outcome of increased inflation and reduced capital intensity is not a necessary implication of a greater deficit. A different debt policy would permit a deficit with no inflation. By totally differentiating equations (21)–(24) with the constraint that $d\pi=0$, it is possible to find the change in the interest rate and corresponding debt policy that permits such a noninflationary increase in the deficit:

$$d\delta=(\pi+n)d(m+b) \tag{31}$$

$$d(m+b)=(fL'-k\beta')di + (Lf'+\beta+k\beta'(1-\tau)f'')dk \tag{32}$$

$$(\sigma(1-\gamma)f'-n)dk=(1-\sigma)nd(m+b) \tag{33}$$

The separate behavior of m and b is irrelevant for determining the change in i that is required to keep the inflation rate unchanged. Equation (33) can be substituted into (32) to eliminate dk and (31) can then be used to eliminate $d(m+b)$. The resulting equation shows that

$$\left.\frac{di}{d\delta}\right|_{d\pi=0}=\left[\sigma(1-\gamma)f'-n - (Lf'+\beta+k\beta'(1-\tau)f'')(1-\sigma)n\right] \div(\pi+n)(\sigma(1-\gamma)f'-n)(fL'-k\beta') \tag{34}$$

The numerator is unambiguously negative since 1) $\sigma(1-\gamma)f'-n<0$ was assumed from stability considerations, and 2) $f'>0$, $\beta'<0$ and $f''<0$ make $-(Lf'+\beta+k\beta'(1-\tau)f'')(1-\sigma)n<0$. The first term of the denominator is positive while, as already noted, the second term is negative. The final term in the denominator is the difference between the effect on the demand for money of an increase in the interest rate on government debt and its effect on the demand for the bonds themselves. Since a higher value of i can be expected to increase the demand for bonds by more than it reduces the demand for money, this term will be taken to be positive. The denominator as a whole is therefore negative. Thus, $di/d\delta>0$.

In short, the interest rate must increase when the deficit increases if the inflation rate is to remain constant. It is easy to understand why this interest rate increase is necessary. Equation (31) indicates that a stable inflation rate requires that $m+b$ must increase with the deficit. This is so because a higher value of $m+b$ permits the larger annual increase in the money supply, and/or government borrowing that must accompany an increased deficit to be absorbed without a higher proportional rate of growth of either money or bonds. To state this same point in a slightly different way, the faster growth of government liabilities can be absorbed without increasing the *proportional* growth rate of either money or bonds if the *level* of money and bonds that is demanded (i.e., the denominators of the proportional growth rates) is increased. The higher interest rate makes this possible by increasing the demand for bonds by more than it decreases the demand for money.

Note that in practice the required change would come about by financing the increased deficit with a higher ratio of bonds to money than had prevailed in the initial equilibrium. Achieving this reduction in liquidity would require paying a higher rate of interest on those bonds.

It is clear that this policy of a higher interest rate and an increased supply of real government debt must reduce the capital intensity of production. This can be seen directly by combining equations (31) and (33):

$$\left.\frac{dk}{d\delta}\right|_{d\pi=0}=\frac{(1-\sigma)n}{(\pi+n)(\sigma(1-\gamma)f'-n)}<0 \tag{35}$$

The cost of avoiding the higher inflation rate that would otherwise accompany an increased deficit is a lower level of capital intensity and a smaller real income.

C. *A Deficit without Reduced Capital Intensity*

The crowding out of real capital accumulation by the government deficit can be avoided by allowing inflation to occur. It is worth examining how much inflation and what change in the interest rate are needed to keep capital intensity unchanged.

Differentiating (24) with $dk=0$ shows immediately that $m+b$ must also remain unchanged. Equation (21) then implies that $d\delta=(m+b)d\pi$, i.e., that an increase in the deficit must increase inflation. With $d(m+b)=0$, equations (22) and (23) together imply

$$0=(fL'-k\beta')di+(1-\tau\lambda)k\beta' d\pi \tag{36}$$

Substituting $(m+b)^{-1}d\delta$ for $d\pi$ yields the change in i required to keep k fixed:

$$\left.\frac{di}{d\delta}\right|_{dk=0}=\frac{-(1-\tau\lambda)k\beta'}{(m+b)(fL'-k\beta')} \tag{37}$$

As explained above, the denominator is positive in the likely case that a rise in the interest rate increases the demand for bonds by more than it decreases the demand for money. The value of λ would be zero if the tax law did not cause inflation to reduce the real net return on capital. In that case, the numerator is positive and $di/d\delta>0$. More generally, even when historic cost depreciation rules do raise the effective tax rate on real profits ($\lambda>0$), the numerator will still be positive as long as inflation raises the *nominal* after-tax return on capital.[19]

The mechanism by which a higher interest rate permits a constant value of k is clear from the derivation. A constant value of k implies a constant value of $m+b$ and therefore an increased value of π. With a constant value of k, a higher rate of inflation would actually decrease b (and therefore $m+b$) unless i is raised to prevent this.

Comparing equation (37) and (34) shows that the increase in i that keeps k unchanged is less than the increase in i that keeps π unchanged:

$$\begin{aligned}\left.\frac{di}{d\delta}\right|_{d\pi=0}-\left.\frac{di}{d\delta}\right|_{dk=0}=&\big[\sigma(1-\gamma)f'-n\\ &-(Lf'+\beta+k\beta'(1-\tau)f'')(1-\sigma)n(m+b)\\ &-(\pi+n)(\sigma(1-\gamma)f'-n)(1-\tau\lambda)k\beta'\big]/\\ &\big[(m+b)(fL'-k\beta')(\pi+n)\\ &\times(\sigma(1-\gamma)f'-n)>0\big]\end{aligned} \tag{38}$$

The reason for this is clear. Holding k and therefore $m+b$ constant implies $d\pi>0$. Making $d\pi=0$ requires an increase in $m+b$ and therefore a higher rate of interest.

D. *A Deficit Financed by Interest-Bearing Debt*

A particularly interesting debt policy requires that any increase in the deficit be financed only by additional borrowing. The real growth of the money supply remains constant. This section looks briefly at the effect of such a policy. This specification of debt policy implies that the real rate of new money creation remains unchanged: DM/pN is constant. Since $DM/pN=m(DM/M)=m(\pi+n)$, this debt policy implies $m(\pi+n)=c$, a constant. The implication of this for capital intensity and inflation depends on the interest elasticity of the demand for money. On the simplifying assumption that the money demand is completely inelastic ($L'=0$), it is easily shown that the deficit unambiguously decreases capital intensity and increases inflation. With $m(\pi+n)=c$,

$$d\delta=(\pi+n)db+bd\pi \tag{39}$$

$$dm=\frac{-m}{\pi+n}d\pi \tag{40}$$

---

[19] The nominal after-tax return on capital is $(1-\tau)f'+(1-\tau\lambda)\pi$.

Combining $m=Lf$ and $m(\pi+n)=c$ yields

$$(\pi+n)f'dk=-fd\pi \tag{41}$$

Thus inflation and capital intensity move in opposite directions. The growth equilibrium condition of equation (24) implies

$$\frac{\sigma(1-\gamma)f'-n}{(1-\sigma)n}dk=db+dm \tag{42}$$

Using (40) and (41) to eliminate dm yields

$$db=\left(\frac{\sigma(1-\gamma)f'-n}{(1-\sigma)n}-\frac{mf'}{f}\right)dk \tag{43}$$

Equation (39) can thus be rewritten using (41) and (43) as

$$d\delta=(\pi+n)\left\{\frac{(1-\gamma)f'-n}{(1-\sigma)n}-\frac{mf'}{f}-\frac{bf'}{f}\right\}dk \tag{44}$$

This shows that $dk/d\delta<0$ and (41) then implies that $d\pi/d\delta>0$. Thus even with a debt policy that keeps the growth of real money balances unchanged, the deficit increases inflation and reduces capital intensity.

III. Fiscal Incentives, Saving, and Aggregate Demand

The three-asset growth model can be used to analyze the effects of an exogenous increase in the saving rate.[20] The most important issue to be examined is the possibility of excess saving. Under quite reasonable conditions, an increase in the saving rate will be absorbed into higher capital intensity without any problem for aggregate demand if there is accommodating monetary policy. The possibility of excess saving arises when investors are unwilling to hold real capital in their portfolio at a lower rate of return; I shall refer to this as a "safety trap" by analogy to the traditional Keynesian liquidity trap. The problem is exacerbated if the yield on government bonds also cannot be lowered, that is, if the economy is also in a liquidity trap.

The problem of excess saving can manifest itself in two ways. Under some conditions, the extra saving could be absorbed in additional capital if the steady-state rate of inflation is reduced. If there is no inflation in the initial equilibrium, the increased saving rate would involve a continuous price deflation. While there may be no theoretical problem with this, as a practical matter the downward rigidity of money wages could prevent this from occurring. The additional saving would not be absorbed but would result in unemployment. The problem is even worse when the safety trap and liquidity trap conditions both prevail. Under these conditions, the extra savings cannot be absorbed in increased capital even if the inflation rate could be permanently reduced.

The problem of excess saving arises only if the government restricts its accommodating action to monetary policy. This section shows how tax incentives, or under some conditions an increased deficit, can be used to assure that an increase in the saving rate results in a greater rate of capital formation.

A. *Increased Saving with Accommodating Monetary Policy*

Before studying the problem of excess saving, it is useful to examine the nature of the well-behaved equilibrium in which additional saving can be absorbed with the help of only accommodating monetary policy. Let us impose the requirement that the real deficit (δ), the share of the government in national income (γ), and the tax rates on capital income (τ and γ) remain constant. The rate of inflation will also be required to remain unchanged, thus precluding the problem of unattainable price deflation.

The key change from the analysis of Section II is that the differential of the growth equilibrium condition (equation (24)) now involves a change in the saving rate:

$$[(1-\gamma)f+n(m+b)]d\sigma + [\sigma(1-\gamma)f'-n]dk-(1-\sigma)nd(m+b)=0 \tag{45}$$

---

[20] Such an increase in the saving rate (σ) may reflect a change in taste or a change in institutions such as Social Security that are not explicitly included in the model.

with a constant deficit ($d\delta=0$) and constant inflation rate ($d\pi=0$), the government budget constraint

$$d\delta=(\pi+n)d(m+b)+(m+b)d\pi \tag{46}$$

implies $d(m+b)=0$. Together with (45) this shows immediately that the higher saving rate increases capital intensity

$$\frac{dk}{d\sigma}=-\frac{(1-\gamma)f+n(m+b)}{\sigma(1-\gamma)f'-n}>0 \tag{47}$$

The required change in the interest rate can then be derived from the money-demand relation (equation (22)) and the bond-demand relation (equation (23)). Together these imply

$$\begin{aligned} d(m+b)&=[Lf'+\beta+k\beta'(1-\tau)f'']dk \\ &\quad+(fL'-k\beta')di+k\beta'(1-\tau\lambda)d\pi \end{aligned} \tag{48}$$

With $d(m+b)=d\pi=0$,

$$\frac{di}{dk}=-\frac{Lf'+\beta+k\beta'(1-\tau)f''}{fL'-k\beta'} \tag{49}$$

If the effect of a change in the interest rate on the demand for bonds exceeds its effect on the demand for money, the denominator is positive and $(di/dk)<0$. To achieve this reduction in the equilibrium interest rate on government debt, the money supply must be increased relative to real income and the supply of bonds must be reduced relative to the capital stock. The precise changes are indicated by equations (22) and (23) and satisfy $d(m+b)=0$.

In short, if investors are willing to accept a lower return on capital accompanied by a less than equal reduction in the yield on government debt, an increase in the saving rate can raise capital intensity without any change in inflation or other government policies.[21]

B. *Safety Preference and Excess Saving*

The basic insight of the Keynesian analysis is that, in a monetary economy, additional saving will not automatically be invested. When the yield on real capital becomes too low, individuals will prefer to hold government bonds rather than to assume the greater risk associated with the ownership of real capital. More precisely, the demand for government bonds becomes infinitely elastic at some low differential between the yield on real capital $((1-\tau)f'+(1-\tau\lambda)\pi)$ and the yield on government bonds (i). In the notation of the bond-demand equation,

$$b=k\beta((1-\tau)f'+(1-\tau\lambda)\pi-i) \tag{50}$$

assume that $\beta'=-\infty$ at some low value of $(1-\tau)f'+(1-\tau\lambda)\pi-i$. We can refer to this situation as being in a safety trap.

When the economy has reached this condition, a further fall in $(1-\tau)f'+(1-\tau\lambda)\pi-i$ is not possible. This additional constraint on the adjustment of the economy can be the source of an excess saving problem. When certain further conditions exist, the increase in capital intensity could only occur if the rate of inflation could be reduced. If the initial equilibrium had no inflation (or a very low rate), the required reduction in inflation might not be achievable and the extra saving would result in an unemployment disequilibrium.[22]

To see the conditions under which this problem would arise, note that the safety trap condition implies

$$(1-\tau)f''dk+(1-\tau\lambda)d\pi-di=0 \tag{51}$$

This in turn implies that $db=\beta dk$ and therefore that

$$d(m+b)=fL'di+(Lf'+\beta)dk \tag{52}$$

With no change in the government deficit, the government budget constraint entails

---

[21] If the interest rate on government debt cannot be reduced, an increase in the rate of inflation could achieve the same thing (as long as $1-\tau\lambda>0$). This is essentially Tobin's solution since he assumes $di=0$. The implication of Section IIIA and the remainder of Section III is that Tobin's inflationary policy is unnecessary and may be counterproductive.

[22] The dynamics of such an employment disequilibrium will not be considered. The relative strength of the Pigou effect and Wicksell effect would influence the ultimate path. For the current purpose, it is sufficient that price deflation and unemployment would be required for at least some period of time.

$$(53)\quad (\pi+n)d(m+b)+(m+b)d\pi=0$$

Using this equation to eliminate $d(m+b)$ from (52) and using (51) to rewrite di in terms of $d\pi$ and dk yields

$$(54)\quad \frac{d\pi}{dk}=-\frac{fL'(1-\tau)f''+Lf'+\beta}{\dfrac{m+b}{\pi+n}+fL'(1-\tau\lambda)}$$

The numerator is unambiguously positive. If the denominator is also positive, an increase in capital intensity must be accompanied by a lower rate of inflation. There are two different plausible conditions under which the denominator will be positive. If the demand for money is interest inelastic ($L'=0$), or, more generally, if the effect of interest on money demand is small ($((\pi+n)fL'(1-\tau\lambda)<m+b)$, the denominator will be positive. Alternatively, regardless of the size of L', if inflation raises the effective tax rate on capital income by enough to make the nominal after-tax yield on capital vary inversely with inflation ($1-\tau\lambda<0$), both terms in the denominator will be positive. In either case, increased capital intensity could not accompany a higher saving rate unless the rate of inflation could be reduced.

It is easy to see why the safety trap condition implies that a greater capital intensity entails a lower rate of inflation. Consider the case of inelastic money demand. The safety trap implies that the demand for bonds increases in proportion to the capital stock: $db=\beta dk$. With inelastic money demand, the money supply also increases with the capital stock: $dM=Lf'dk$. But with no change in the government deficit, the steady-state value of $m+b$ can increase only if the inflation rate is lower. In the alternate case in which inflation increases the effective tax rate, the analysis is only slightly more complex. If there were no change in inflation, the interest rate would have to fall to maintain the minimum yield differential with greater capital intensity. But this would increase the demand for money, raising $m+b$. This is incompatible with a constant inflation rate. If the inflation rate increased, this would further reduce the yield on capital relative to government bonds.

The problem of excess saving under these conditions can be avoided if the government uses fiscal policy as well as monetary policy to accommodate the additional saving. Consider first the possibility of responding to a higher saving rate by reducing the rate of tax on capital income, τ. With a lower tax rate, the net of tax yield on the real capital stock can be maintained while the greater capital intensity depresses the pretax yield. To confirm explicitly that this fiscal incentive is sufficient to permit greater capital intensity, with no change in the rate of inflation, consider the four equations that describe the safety trap equilibrium with $d\tau\neq 0$. The government budget constraint with $d\pi=0$ and $d\delta=0$ implies $d(m+b)=0$. Substituting this into the growth equilibrium (equation (45)) shows $dk/d\sigma>0$ exactly as in equation (47). The two remaining conditions that must be satisfied are the safety trap condition with $d\pi=0$:

$$(55)\quad (1-\tau)f''dk-di-(f'+\lambda\pi)d\tau=0$$

and the condition that the change in the demands for debt and money leave $m+b$ unchanged:

$$(56)\qquad 0=fL'di+(\beta+Lf')dk$$

Equation (56) shows that the interest rate must rise, reducing the demand for money per unit of capital. With the unique increase in di determined by (56) and the unique increase in k determined by (47), equation (55) shows the required decrease in the tax rate τ.

A higher saving rate can be transformed into greater capital intensity with no change in inflation even without changing the tax rate on capital income by an accommodating increase in the government deficit accompanied by a lower rate of interest. The lower rate of interest balances the fall in the return on real capital, permitting the real capital to be absorbed. The greater deficit with the unchanged rate of inflation permits an increase in both money and bonds that is required by the fall in i and increase in k.

To see all of this explicitly, note that with $d\tau = d\pi = 0$, the safety trap condition becomes

$$(57)\qquad (1-\tau)f''dk - di = 0$$

The increased demand for money and bonds is

$$(58)\qquad d(m+b) = fL'di + (Lf' + \beta)dk$$

which, from (57), is

$$(59)\qquad d(m+b) = (fL'(1-\tau)f'' + Lf' + \beta)dk$$

Substituting this value of $d(m+b)$ into the growth equilibrium (equation (45)) yields

$$(60)\qquad \frac{dk}{d\sigma} = -\frac{(1-\gamma)f + n(m+b)}{(1-\gamma)f' - n - (fL'(1-\tau)f'' + Lf' + \beta)}$$

This is unambiguously positive. The last term in the denominator, which reflects the increased deficit (i.e., $d(m+b)$), reduces the size of $dk/d\sigma$ but does not alter the fact that it is positive. With dk determined by (60), equation (59) implies a unique value of $d(m+b) > 0$. The government budget constraint with $d\pi = 0$ then gives the required change in the deficit, $d\delta = (\pi + n)\, d(m+b) > 0$.

Although both the reduced tax on capital income and the increased deficit are capable of turning additional saving into greater capital intensity without a change in the price level (and therefore without the possibility of a deflationary unemployment disequilibrium), the reduced tax on capital income has at least three advantages over the increased government deficit. First, and probably most important, the equilibrium capital intensity is greater if the increased saving is accommodated by a lower tax rate. Second, the lower tax rate on capital income reduces the excess burden caused by a distorting tax.[23] Finally, the tax reduction can be effective under special conditions when the increased deficit would fail. More specifically, the lower interest rate that must accompany the increased deficit would not be possible if the economy is also in a liquidity trap, that is, if investors are unwilling to hold any asset other than money at a lower rate of interest.[24] Even in this case, the tax reduction (and increased rate of interest) can be used to accommodate a higher rate of saving.

IV. Some Conclusions

This paper has studied the long-run impact of fiscal policies on inflation and capital formation. The analysis uses an expanded monetary growth model in which the government finances its deficit by issuing both money and interest-bearing debt.

One major focus of the paper is the effect of a permanent increase in the government's real deficit in a fully employed economy. An important conclusion is that such an increased deficit must raise the rate of inflation or lower the capital intensity of production or both. The analysis shows that the combination of *both* increased inflation and reduced capital intensity is a likely outcome with current *U.S.* tax rules and the prevailing monetary policy of allowing the interest rate to rise with inflation in a way that keeps the real interest rate unchanged. Section II determines the debt management policy (and the corresponding change in the interest rate) that would be required to maintain either a constant inflation rate or a constant capital-labor ratio.

---

[23] The welfare gain from reducing the tax rate on capital income depends on the way in which the lost tax revenue is recovered. My 1978b paper shows that even when the uncompensated elasticity of saving with respect to its return is zero, the excess burden of the tax system would be reduced by lowering the tax on capital income and raising it on labor income. In the current context there is the further complication that the increased deficit and lower interest rate would permit lower total taxes.

[24] In the notation of the model, $L' = -\infty$ at some low level of i. This implies the extra constraint $di \geqslant 0$ which is consistent with equation (56) in the context of a tax rate reduction but not with equation (57) when $d\tau = 0$ and $d\delta > 0$. An increased deficit could avoid unemployment by the Keynesian remedy of absorbing all of the additional saving, i. e., with $dk = 0$.

The second purpose of the paper is to analyze the effect of an exogenous increase in the saving rate and the possibility of "excessive saving." The problem of excessive saving arises when the yield on capital becomes so low that individuals prefer to hold government bonds rather than the more risky claims to real capital. Under some conditions, an increased rate of saving could only be absorbed in increased capital intensity if the rate of inflation could be permanently reduced. This requirement might entail a negative inflation rate which, as a practical matter, would be precluded by the downward rigidity of money wages. In this case, the additional saving would not be absorbed but would result in unemployment.

Section III shows first that there is no problem of excessive saving if 1) investors are willing to hold real capital even though the differential between its yield and that on government bonds is narrowed, and if 2) the government reduces the interest rate on government bonds by expanding the money supply more rapidly than the stock of bonds. When these conditions are met, an increase in saving can be absorbed in greater capital intensity without any change in either inflation or the government deficit.

A problem can arise if the economy is in a safety trap, that is, if investors would be unwilling to hold real capital if the difference between its yield and that on government debt were reduced. In that case, an increased saving rate can imply price deflation and therefore possible unemployment. This problem can be avoided however by reducing the tax on capital income (or, in some cases, by an increased deficit that absorbs some but not all of the higher savings rate). In short, by using fiscal incentives as well as monetary accommodation, an increased saving rate can be converted to greater capital intensity.

The analysis as a whole, although clearly a theoretical study of a simplified economy, suggests some insights that may help in understanding the unsatisfactory macroeconomic experience of the past decade and in designing more appropriate economic policies for the future. The recent years have been characterized by substantial inflation, a low rate of investment, and large government deficits. Section II shows how an increased government deficit can give rise to both greater inflation and reduced capital intensity. The combination of inflation and historic cost depreciation raised the effective tax rate on the income from real capital while the monetary and debt management policies have kept the real interest rate on government debt unchanged. The reduced-equilibrium capital-labor ratio that this implies manifests itself as a lower rate of investment. The problem is then exacerbated when the government responds to decreased investment by further enlarging its deficit. The analysis suggests that a more appropriate solution would be to reduce the deficit while stimulating investment through a lower tax rate and a depreciation method based on current rather than historic costs.

I have argued elsewhere that the United States should increase its saving rate to take advantage of the high social rate of return on additional investment.[25] Such an increase in the private saving rate could be achieved by reducing the growth of Social Security benefits or by tax reforms that make the personal income tax more like a consumption tax. These proposals implicitly assumed that such extra saving would result in greater capital intensity rather than in a fall in aggregate demand. Section III implies that this assumption is warranted. With appropriate fiscal inventives and accommodating monetary policies, an increase in saving can be absorbed in greater capital intensity without any change in the rate of inflation.

---

[25] See my 1976 and 1977 papers.

REFERENCES

W. J. Baumol, "The Transactions Demand for Cash: an Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

O. Eckstein and R. Brinner "The Inflation Process in the United States," study for the Joint Economic Comm., 92d Cong., 2d sess. 1972.

M. S. Feldstein, "Inflation, Income Taxes and the Rate of Interest: A Theoretical Analysis," *Amer. Econ. Rev.*, Dec. 1976, *66*, 809–20.

______, "Does the United States Save Too Little?," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 116–21.

______, "Inflation and the Stock Market," *Amer. Econ. Rev.*, Dec. 1980 forthcoming.

______, (1978a) "The Effect of Inflation on the Prices of Land and Gold," Nat. Bur. Econ. Res. work. paper no. 296, Cambridge, Mass., Nov. 1978.

______, (1978b) "The Welfare Cost of Capital Income Taxation," *J. Polit. Econ.*, Apr. 1978, *86*, Part II, S2–S51.

______, "National Savings in the United States," in Eli Shapiro and William L. White, eds., *Investment and Saving for Productivity, Growth and High Employment*, Englewood Cliffs, 1977.

______ **and G. Chamberlain,** "Multimarket Expectations and the Rate of Interest," *J. Money, Credit, Banking*, Nov. 1973, *5*, 873–902.

______ **and O. Eckstein,** "The Fundamental Determinants of the Interest Rate," *Rev. Econ. Statist.*, Nov. 1970, *52*, 363–75.

______ **and J. S. Flemming,** "The Problem of Time-Stream Evaluation," *Bull. Oxford Univ. Inst. Econ. Statist.*, Feb. 1964, *26*, 79–85.

______, **J. Green, and E. Sheshinski,** "Corporate Financial Policy and Taxation in a Growing Economy," *Quart. J. Econ.*, Aug. 1979, *93*, 411–32.

______, ______, **and** ______, "Inflation and Taxes in a Growing Economy with Debt and Equity Finance," *J. Polit. Econ.*, Apr. 1978, *86*, Part II, S53–S70.

______ **and M. Rothschild,** "Towards an Economic Theory of Replacement Investment," *Econometrica*, May 1974, *42*, 393–423.

______ **and L. Summers,** "The Rate of Profit: Falling or Cyclical?," *Brookings Papers*, Washington 1977, *1*, 211–28.

______ **and** ______, "Inflation and the Taxation of Capital Income in the Corporate Sector," *Nat. Tax J.*, Dec. 1979, *32*, 445–70.

Irving Fisher, *The Theory of Interest*, New York 1954.

B. J. Friedman, "Financial Flow Variables and the Short-Run Determination of Long-Term Interest Rates," *J. Polit. Econ.*, Aug. 1977, *85*, 661–89.

______, "Crowding Out or Crowding In? Economic Consequences of Financial Government Deficits," *Brookings Papers*, Washington 1978, *3*, 593–654.

R. J. Gordon "The Incidence of the Corporation Income Tax in U.S. Manufacturing," *Amer. Econ. Rev.*, Sept. 1967, *57*, 731–58.

______, "Inflation in Recession and Recovery," *Brookings Papers*, Washington 1971, *1*, 105–66.

J. Green and E. Sheshinski, "Budget Displacement Effects of Inflationary Finance," *Amer. Econ. Rev.*, Sept. 1977, *67*, 671–82.

A. C. Harberger, "The Incidence of the Corporate Income Tax," *J. Polit. Econ.*, June 1962, *70*, 215–40.

Seymour Harris, *Economics of Social Security*, New York 1941.

R. F. Harrod, *Towards a Dynamic Economics*, London 1948.

John Maynard Keynes, *The General Theory of Employment, Interest and Money*, London 1936.

Marian Krzyzaniak and Richard Musgrave, *The Shifting of the Corporation Income Tax*, Baltimore 1963.

D. Levhari and D. Patinkin, "The Role of Money in a Simple Growth Model," *Amer. Econ. Rev.*, Sept. 1968, *58*, 713–53.

W. Oakland, "Corporate Earnings and Tax Shifting in U.S. Manufacturing, 1930–1968," *Rev. Econ. Statist.*, Aug. 1972, *54*, 235–44.

T. J. Sargent, "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," *Brookings Papers*, Washington 1973, *2*, 429–72.

______, "Interest Rates and Expected Inflation: A Selective Summary of Recent Research," Nat. Bur. Econ. Res. *Explor. in Econ. Res.*, No. 3, *3*, Summer 1976, 303–25.

J. B. Shoven and J. Whalley, "A General Equilibrium Calculation of the Effects of Differential Taxation of Income from Capital in the U.S.," *J. Publ. Econ.*, Nov. 1972, *1*, 281–321.

J. Stiglitz, "The Corporation Income Tax," *J. Publ. Econ.*, Apr./May 1976, *5*, 303–11.

______, "Taxation, Corporate Financial Policy and the Cost of Capital," *J. Publ. Econ.*, Feb. 1973, *2*, 1–34.

J. Tobin, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–84.

______ **and W. Buiter,** "Fiscal and Monetary Policies, Capital Formation and Economic Activity," in George von Furstenberg, ed., *The Government and Capital Formation*, forthcoming.

W. P. Yohe and D. S. Karnovsky, "Interest Rates and Price Level Changes, 1952–69," *Rev. Fed. Reserve Bank St. Louis*, Dec. 1969, *51*, 18–38.

A Model of Sales

By Hal R. Varian*

Economists have belatedly come to recognize that the "law of one price" is no law at all. Most retail markets are instead characterized by a rather large degree of price dispersion. The challenge to economic theory is to describe how such price dispersion can persist in markets where at least some consumers behave in a rational manner. Starting with the seminal paper of George Stigler, a number of economic theorists have proposed models to describe this phenomenon of equilibrium price dispersion. See, for example, Gerard Butters, John Pratt, David Wise, and Richard Zeckhauser, Michael Rothschild, Steven Salop, Salop and Joseph Stiglitz (1977), Yuval Shilony, Stiglitz, and Louis Wilde and Alan Schwartz.

Most of the models of price dispersion referred to above are concerned with analyzing "spatial" price dispersion; that is, a situation where several stores contemporaneously offer an identical item at different prices. A nice example of such a model is the "bargains and ripoffs" paper of Salop and Stiglitz (1977). They consider a market with two kinds of consumers; the "informed" consumers know the entire distribution of offered prices, while the "uninformed" consumers know nothing about the distribution of prices. Hence the informed consumers always go to a low-priced store, while the uninformed consumers shop at random. The stores have identical U-shaped cost curves and behave as monopolistically competitive price setters. Salop and Stiglitz show that for some parameter configurations, the market equilibrium takes a form where some fraction of the stores sell at the competitive price (minimum average cost) and some fraction sell at a higher price. The high-price stores' clientèle consists only of uninformed consumers, but there is a sufficiently large number of them to keep the stores in business.

In the Salop and Stiglitz model—as in all the models of spatial price dispersion—some stores are supposed to *persistently* sell their product at a lower price than other stores. If consumers can learn from experience, this persistence of price dispersion seems rather implausible.

An alternative type of price dispersion might be called "temporal" price dispersion. In a market exhibiting temporal price dispersion, we would see each store varying its price over time. At any moment, a cross section of the market would exhibit price dispersion; but because of the intentional fluctuations in price, consumers cannot learn by experience about stores that consistently have low prices, and hence price dispersion may be expected to persist.

One does not have to look far to find the real world analog of such behavior. It is common to observe retail markets where stores deliberately change their prices over time—that is, where stores have *sales*. A casual glance at the daily newspaper indicates that such behavior is very common. A high percentage of advertising seems to be directed at informing people of limited duration sales of food, clothing, and appliances.

Given the prevalence of sales as a form of retailing, it is surprising that so little attention has been paid to sales in the literature of economic theory. In fact, I know of no work in economic theory that *explicitly* examines the rationale of price dispersion by means of sales.[1] However, the work of Shilony does provide an implicit rationale for the use of sales as a marketing device.

---

*University of Michigan. Research support by the National Science Foundation and the Guggenheim Memorial Foundation is gratefully acknowledged. Helpful comments were received from James Adams, Paul Courant, Vincent Crawford, and John Panzar. Research assistance was provided by Todd Lanski.

[1] Salop and Stiglitz' 1976 paper is concerned with "spatial" price dispersion rather than temporal price dispersion.

Shilony examines an oligopolistic market where consumers can purchase costlessly from neighborhood stores, but incur a "search cost" if they venture to more distant stores in search of a lower price. He shows that no Nash equilibrium exists in pure pricing strategies. On the other hand, Shilony does establish the existence of an equilibrium *mixed* strategy—that is, a strategy where firms randomize their prices. Such a strategy could be interpreted as stores having randomly chosen sales.

In this paper, I explicitly address the question of sales equilibria. The model may be regarded as a combination of the Salop-Stiglitz and the Shilony models described above. As in the Salop-Stiglitz model, it will be assumed that there are informed and uninformed consumers. As in the Shilony model, I will allow for the possibility of randomized pricing strategies by stores. I will be interested in characterizing the equilibrium behavior in such markets.

In the model to be described below, firms engage in sales behavior in an attempt to price discriminate between informed and uninformed customers. This is of course only one aspect of real world sales behavior. Other reasons for sales behavior might include inventory costs, cyclical fluctuations in costs or demand, loss leader behavior, advertising behavior, and so on. The theoretical examination of these motives is left for future work.

I. The Model

Let us suppose there is a large number of consumers who each desire to purchase, at most, one unit of some good. The maximum price any consumer will pay for the good—a consumer's reservation price—will be denoted by r. Consumers come in two types, informed and uninformed.[2] Uninformed consumers shop for the item by choosing a store at random; if the price of the item in that store is less than r, the consumer purchases it. Informed consumers, on the other hand, know the whole distribution of prices, and in particular they know the lowest available price at any time. Hence, they go to the store with the lowest price and purchase the item there.

One might think of a model where stores advertise their sale prices in the weekly newspaper. Informed consumers read the newspaper and uninformed consumers do not. Let $I>0$ be the number of informed consumers, and $M>0$ the number of uninformed consumers. Let n be the number of stores, and let $U=M/n$ be the number of uninformed consumers per store.

Each store has a density function $f(p)$ which indicates the probability with which it charges each price p. In its choice of this pricing strategy, each firm takes as given the pricing strategies chosen by the other firms and the demand behavior of the consumers. Only the case of a symmetric equilibrium will be examined, where each firm chooses the same pricing strategy.[3]

Each week, each store randomly chooses a price according to its density function $f(p)$. A store *succeeds* in its sale if it turns out to have the lowest price of the n prices being offered. In this case the store will get $I+U$ customers. If a store *fails* to have the lowest price, it will get only its share of uninformed customers, namely U. If two or more stores charge the lowest price, it will be considered a *tie*, and the low-price stores will each get an equal share of the informed customers.

Finally the stores are characterized by identical, strictly declining average cost curves.[4] The cost curve of a representative firm will be denoted by $c(q)$. It will be assumed that entry occurs until (expected) profits are driven to zero. Thus we will be examining a symmetric monopolistically competitive equilibrium in pricing strategies.

---

[2] For now, the uninformed-informed distinction is exogenously given. The decision to become informed or uninformed will be examined in Section III.

[3] Some justification for this symmetry assumption is given by Proposition 9 in the Appendix.

[4] The motivation for this assumption is the casual observation that retail stores are characterized by fixed costs of rent and sales force, plus constant variable costs—the wholesale cost—of the item being sold.

II. The Analysis

The maximum number of customers a store can get is $I+U$. Let $p^*=c(I+U)/(I+U)$ be the average cost associated with this number of customers.

PROPOSITION 1: *$f(p)=0$ for $p>r$ or $p<p^*$.*

PROOF:

No price above the reservation price will be charged since there is zero demand at any such price. No price less that p^* will be charged since only negative profits can result from such a price.

PROPOSITION 2: *There is no symmetric equilibrium where all stores charge the same price.*

PROOF:

Suppose that all stores were charging a single price p with $r \geqq p > p^*$. Then a slight cut in price by one of the stores would capture all of the informed market, and thus make a positive profit. If all stores were charging p^*, each would get an equal share of the market and thus be making negative profits.

Proposition 2 is simply a variant of the well-known argument that declining average cost curves and "competitive" behavior are incompatible. I therefore concentrate on establishing the nature of a price-randomizing solution. Recall that p is a *point mass* of a probability density function f if there is positive probability concentrated at p.

PROPOSITION 3. *There are no point masses in the equilibrium pricing strategies.*[5]

PROOF:

The intuition of this argument is seen to be quite straightforward. If some price p were charged with positive probability, there would be a positive probability of a tie at p. If a deviant store charged a slightly lower price, $p-\varepsilon$, with the same probability with which the other stores charged p, it would lose profits on order ε, but gain a fixed positive amount of profits when the other stores tied. Thus for small ε its profits would be positive, contradicting the assumption of equilibrium.

Let us proceed to a detailed formulation of this argument. First note that p^* can never be charged with positive probability, for when p^* is the lowest price charged, profits are zero, and if there is a tie at p^*, profits are negative. Suppose then that $p>p^*$ is charged with positive probability.

The number of points of positive mass in any probability distribution must be countable so we can find an arbitrarily small ε such that $p-\varepsilon$ is charged with probability 0. Consider what happens if we charge $p-\varepsilon$ with the probability with which we used to charge p, and charge p with probability 0. The increase in profits will be

$$\begin{aligned}
&Pr(P_i>p-\varepsilon \text{ all } i,\ P_i\neq p \text{ any } i)\\
&\qquad ((p-\varepsilon)(I+U)-c(I+U))\\
&\quad -Pr(P_i>p \text{ all } i)\ (p(I+U)-c(I+U))\\
&\quad +Pr(p_i<p-\varepsilon \text{ some } i)\quad ((p-\varepsilon)U-c(U))\\
&\quad -Pr(P_i<p \text{ some } i)(pU-c(U))\\
&\quad +\sum_{k=2}^{n} Pr(P_i\geqq p-\varepsilon \text{ all } i,\ P_i=p \text{ for } k \text{ stores})\\
&\qquad ((p-\varepsilon)(I+U)-c(I+U))\\
&\quad -\sum_{k=2}^{n} Pr(P_i\geqq p \text{ all } i,\ P_i=p \text{ for } k \text{ stores})\\
&\qquad (p(U+I/k)-c(U+I/k))
\end{aligned}$$

As ε approaches zero, the sum of the first four terms approaches zero, while the sum of the last two terms remains a positive number. Hence for small ε profits are positive, contradicting the assumption of an equilibrium strategy.

Proposition 3 expresses the essential difference between models of spatial price dispersion and models of temporal price dispersion. Most models of spatial price dispersion, such as the Salop-Stiglitz model or the Wilde-Schwartz model, have equilibria with specific prices being charged with positive probability mass. The above argument shows that such strategies cannot be profit-maximizing Nash behavior in a temporal randomizing model.

---

[5] Proposition 9 in the Appendix provides a partial converse to this assertion.

Since there are no point masses in the equilibrium density, the cumulative distribution function will be a continuous function on (p^*, r). Let $F(p)$ be the cumulative distribution function for $f(p)$; thus $f(p)=F'(p)$ almost everywhere.

We can now construct the expected profit function for a representative store. When a store charges price p, exactly two events are relevant. It may be that p is the smallest price being charged, in which case, the given store gets all of the informed customers. This event happens only if all the other stores charge prices higher than p, an event which has probability $(1-F(p))^{n-1}$. On the other hand, there may be some store with a lower price, in which case the store in question only gets its share of the uninformed customers. This event happens with probability $1-(1-F(p))^{n-1}$. (By Proposition 3 we can neglect the probability of any ties.) Hence the expected profit of a representative store is

$$\int_{p^*}^{r}\{\pi_s(p)(1-F(p))^{n-1} + \pi_f(p)[1-(1-F(p))^{n-1}]\}f(p)\,dp$$

where $\pi_s(p)=p(U+I)-c(U+I)$

$$\pi_f(p)=pU-c(U)$$

The maximization problem of the firm is to choose the density function $f(p)$ so as to maximize expected profits subject to the constraints:

$$f(p)\geqq 0;\quad \int_{p^*}^{r} f(p)\,dp=1$$

It is clear that all prices that are charged with positive density must yield the same expected profit; for if some price yields a greater profit than some other price it would pay to increase the frequency with which the more profitable price were charged. Since we require zero profits due to free entry, this common level of profit must be zero.[6] This argument yields

PROPOSITION 4: *If* $f(p)>0$, *then*

$$\pi_s(p)(1-F(p))^{n-1} + \pi_f(p)[1-(1-F(p))^{n-1}]=0$$

(Of course, Proposition 4 also follows directly from the application of the Kuhn-Tucker theorem to the specified maximization problem.) Rearranging this equation, we have a formula for the equilibrium cumulative distribution function:

$$1-F(p)=\left(\frac{\pi_f(p)}{\pi_f(p)-\pi_s(p)}\right)^{\frac{1}{n-1}}$$

Note that the denominator of this fraction is negative for any p between p^* and r. Hence the numerator must be negative so that profits in the event of failure are definitely negative. The construction of $(1-F(p))^{n-1}$ is illustrated in Figure 1. At each p where $f(p)>0$ we can construct $\pi_f(p)$ and $\pi_s(p)$ as illustrated and take the relevant ratio. Proposition 4 gives us an explicit expression for the equilibrium distribution function at those values of p where $f(p)>0$. If this is to be a legitimate candidate for a cumulative distribution function, it should be an increasing function of p. This is easy to verify:

PROPOSITION 5: $\pi_f(p)/(\pi_f(p)-\pi_s(p))$ *is strictly decreasing in p.*

PROOF:

Taking the derivative it suffices to show that

$$(\pi_f(p)-\pi_s(p))U-\pi_f(p)(-I)<0$$

---

[6] One can also formulate the model with a fixed number of firms. In this case, expected profits must be equal to $\Pi_f(r)$.

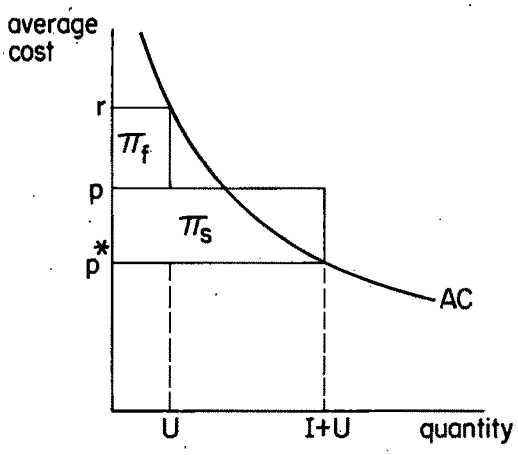


FIGURE 1. GRAPHICAL DETERMINATION OF $\pi_s(p)$ AND $\pi_f(p)$

Using the definitions of π_f and π_s, this can be rearranged to yield

$$\frac{c(I+U)}{I+U}<\frac{c(U)}{U}$$

which is obvious since average cost has been assumed to strictly decrease.

Of course, Proposition 4 characterizes the equilibrium density function only for those prices where $f(p)>0$. In order to fully characterize the equilibrium behavior, we need to establish which prices are charged with positive density.

First, it is clear that prices close to p^* must be charged with positive density:

PROPOSITION 6: *$F(p^*+\varepsilon)>0$ for any $\varepsilon>0$.*

PROOF:

If not, some store could charge $p^*+\varepsilon/2$, and thereby undercut the rest of the market and make positive profits.

Similarly we can characterize the behavior of $f(p)$ near its upper limit.

PROPOSITION 7: *$F(r-\varepsilon)<1$ for any $\varepsilon>0$.*

PROOF:[7]

Suppose not, and let $\hat{p}<r$ be the highest price that is ever charged so that $F(\hat{p})=1$. When $\hat{p}$ is charged, the store will only get the uninformed customers since with probability 1 some other store will be charging a lower price. Since the store must get zero expected profits at each price charged, we must have $\hat{p}U-c(U)=0$. But then $rU-c(U)>0$, so charging r with probability 1 could make a positive profit.

Propositions 6 and 7 show that prices near p^* and r are charged with positive density. It is now easy to show:

PROPOSITION 8: *There is no gap (p_1,p_2) where $f(p)\equiv 0$.*

PROOF:

If not, let $p_1<\hat{p}<p_2$. Now $\hat{p}$ succeeds in being the lowest price in exactly the same circumstances that p_1 succeeds in being the lowest price; namely, when all other prices are greater than p_2. Similarly, $\hat{p}$ fails to be the lowest price when some store charges a price less than p_1, in which case p_1 also fails to be the lowest price. But in each circumstance, since $\hat{p}>p_1$, $\hat{p}$ will make larger profits than p_1. Since p_1 must make zero profits, this shows that charging $\hat{p}$ with probability 1 will make positive profits.

We now have a complete characterization of the equilibrium density: $f(p)>0$ for all p in (p^*,r) and $f(p)=F'(p)$, where

$$(1)\qquad F(p)=1-\left(\frac{\pi_f(p)}{\pi_f(p)-\pi_s(p)}\right)^{\frac{1}{n-1}}$$

We can also solve for the endogenous variables n and p^*. First, note that if a store charges r, it only gets the uninformed customers, and profits must therefore satisfy $\pi_f(r)=0$. Similarly, if a store charges p^* it gets all the informed customers with probability 1 so $\pi_s(p^*)=0$. These two equations can be used to determine n and p^*.

---

[7] A heuristic proof is presented here and a more rigorous proof in the Appendix. (The same holds true for Proposition 8.)

As an example, let us compute the equilibrium density when the cost function has fixed cost $k>0$ and zero marginal cost. Then

$$\pi_s(p)=p(I+U)-k \tag{2}$$

$$\pi_f(p)=pU-k \tag{3}$$

Since $\pi_f(r)=0$, and $U=M/n$

$$rM/n-k=0 \tag{4}$$

or

$$n=rM/k \tag{5}$$

Thus

$$U=M/n=k/r \tag{6}$$

Since $\pi_s(p^*)=0$, we have

$$p^*\left(I+\frac{k}{r}\right)-k=0$$

or

$$p^*=\frac{k}{I+k/r} \tag{7}$$

The equilibrium distribution function can be found by substituting (2) and (3) into (1). We have

$$F(p)=1-\left(\frac{k-pU}{pI}\right)^{\frac{1}{n-1}} \tag{8}$$

Substituting from (6) and rearranging, we find

$$F(p)=1-[(k/I)(1/p-1/r)]^{\frac{1}{n-1}} \tag{9}$$

The equilibrium density function is found by differentiating (9):

$$f(p)=F'(p) = \frac{(k/I)^{\frac{1}{n-1}}}{n-1}\,\frac{(1/p-1/r)^{\frac{1}{n-1}-1}}{p^2} \tag{10}$$

Let

$$m=1-\frac{1}{n-1}=\frac{n-2}{n-1}=\frac{rM-2k}{rM-k} \tag{11}$$

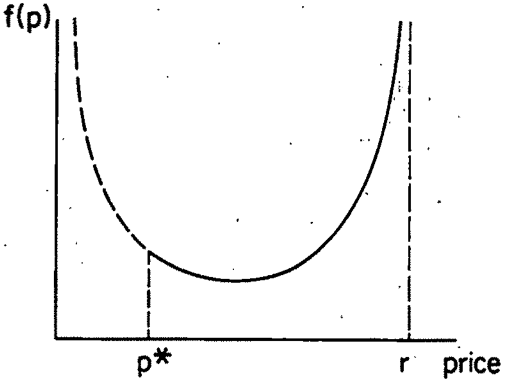


FIGURE 2. GRAPH OF $f(p)=1/p(1-p/r)$

Then $f(p)$ can be written as

$$f(p)=\frac{k(k/I)^{1-m}}{(rM-k)}\,\frac{1}{p^{2-m}(1-p/r)^m} \tag{12}$$

If n is reasonably large, m will be approximately 1, so $f(p)$ will be proportional to

$$\frac{1}{p(1-p/r)} \tag{13}$$

This density is illustrated in Figure 2. Note that stores tend to charge extreme prices with higher probability than they charge intermediate prices. This seems intuitively plausible; a store would like to discriminate in its pricing and charge informed customers p^* (to keep their business) and charge uninformed customers r (to exploit their surplus). Since they are required to sell to all consumers at the same price this tendency shows up in the U-shaped density of prices.

The relevant part of the density is of course that part between p^* and r. Referring to equation (7) we see that if fixed costs are small, or the number of informed consumers is large, p^* will be small. Hence low prices will be charged a high percentage of the time. In some sense the market will be more competitive. On the other hand, the influence of the uninformed consumers is never entirely absent, since high prices will always be charged for some fraction of the time.

TABLE 1—SUMMARY OF COMPARATIVE STATICS

	k	I	M	r	λ
p^*	+	−	0	+	−
n	−	0	+	+	−
$F(p)$	?	+	−	−	+
$\bar{p}$	?	−	+	+	−
$\bar{p}_{min}$	?	?	−	?	?

Note: K=fixed costs, I=number of informed consumers, M=number of uninformed consumers, r=reservation prices, and λ=fraction of informed consumers: $\lambda=I/(I+M)$.

It is of interest to calculate the average price paid by the informed and the uninformed consumers under the equilibrium price density. The average price the uninformed consumers pay is simply:

$$\bar{p}=\int_{p^*}^{r} pf(p)\,dp$$

If we integrate this by parts we have

$$\bar{p}=r-\int_{p^*}^{r} F(p)\,dp$$

Substituting from (9)

$$\bar{p}=p^*+\left(\frac{k}{I}\right)^{\frac{1}{n-1}}\int_{p^*}^{r}\left(\frac{1}{p}-\frac{1}{r}\right)^{\frac{1}{n-1}}dp \tag{14}$$

There seems to be no simple expression for the integral in (14). However in the duopoly case (where $n=2$) the expression becomes rather trivial. Suppose for example that we have $r=1$, $M=2$, $I=1$, $k=1$. Then by (5) we have $n=2$, and by (7), $p^*=.5$. Substituting into (14),

$$\bar{p}=.5\int_{.5}^{1}\left(\frac{1}{p}-1\right)dp=-\ln .5\simeq .69 \tag{15}$$

Let us denote the price paid by the informed customers by p_{min}. Then the density function of p_{min} is just

$$f_{min}(p)=(1-F(p))^{n-1}f(p)$$

Substituting from (9)

$$f_{min}(p)=\frac{k}{I}(1/p-1/r)f(p)$$

Thus, $$\bar{p}_{min}=\frac{k}{I}\int_{p^*}^{r} p(1/p-1/r)f(p)\,dp$$

$$=\frac{k}{I}\left(1-\frac{\bar{p}}{r}\right) \tag{16}$$

For the example given above, $\bar{p}_{min}\simeq .31$.

Table 1 presents the results of the comparative statics computations. The signs are for the most part as expected. However, the behavior of $\bar{p}$ with respect to M—the number of uninformed consumers—does contain an interesting feature. It can be shown that $\partial\bar{p}/\partial M>0$. That is, more uninformed consumers cause the average price paid by the uninformed consumers to rise. This is an example of the detrimental externalities that noneconomizing behavior can impose. However, note from (16) that $\bar{p}_{min}$ will *decrease* with M—the uninformed consumers confer a *beneficial* externality on the informed consumers. This effect seems to arise because the number of stores is increasing in the number of uninformed consumers; since the informed consumers buy at the *lowest* advertised price, a larger number of stores tends to lower the average price they pay.

III. Does it Pay to be Informed?

The model presented above takes the informed and uninformed consumers as exogenously given. However, the decision to become informed or uninformed can easily be made endogenous. Following the Salop-Stiglitz example, let us now suppose that it is possible to become fully informed about the available prices in the market by paying a fixed cost c. We think of this as the cost involved in reading newspaper advertisements, processing the information, and so on. Further, suppose that there are two types of consumers: one group has "search costs" c_2 and the other has search costs c_1, with $c_2 \geqq c_1$.

The decision to be informed or uninformed now depends on the "full price" one pays to purchase the product in question. An uninformed person pays $\bar{p}$ on the average while an informed person pays $\bar{p}_{min}+c_i$, $i=1,2$. In order for the equilibrium to be a full equilibrium, neither the informed, nor the uninformed group should find it in their interest to change their behavior.

For example, suppose $c_2 > c_1 = 0$. Then the low-cost consumers will always be informed. If the high-cost customers also find it in their interest to be informed no equilibrium will exist.[8] It is in the interest of the high-cost consumers to remain uninformed if $\bar{p} < \bar{p}_{min} + c_2$. Using the results of (15) and (16), this reduces to $\bar{p} < k(1-\bar{p}/r)/I + c_2$, or $\bar{p}/r < (k+c_2 I)/(k+rI)$. If c_2 is greater than r, for example, this condition will certainly be satisfied.

IV. Summary

I have shown how stores may find it in their interest to randomize prices in an attempt to price discriminate between informed and uninformed consumers, and have solved explicitly for the resulting monopolistically competitive equilibrium in randomized pricing strategies. The form of the resulting pricing strategy as given in Figure 2 does not seem out of line with commonly observed retailing behavior. Large retailing chains such as Sears and Roebuck and Montgomery Ward sell appliances at their regular price much of the time, but often have sales when the price is reduced by as much as 25 percent. However, we rarely observe them selling an appliance at an intermediate price. Although this casual empiricism can hardly be conclusive, it suggests that the features of the model described here may have some relevance in explaining real world retailing behavior.[9]

APPENDIX

PROPOSITION 7: *$F(r-\varepsilon) < 1$ for any $\varepsilon > 0$.*

PROOF:

If not, let $\hat{p} < r$ satisfy $F(\hat{p}) = 1$ and let (p^i) be a sequence of prices with $f(p^i) > 0$ and $(p^i) \to \hat{p}$. Clearly $F(p^i) \to 1$, so

$$1 = 1 - \left(\frac{\pi_f(\hat{p})}{\pi_f(\hat{p}) - \pi_s(\hat{p})} \right)^{\frac{1}{n-1}}$$

Hence $\pi_f(\hat{p}) = 0$, or $\hat{p}U - c(U) = 0$. But then $rU - c(U) > 0$, so a store charging r with probability 1 would make a positive profit.

PROPOSITION 8: *There is no gap (p_1, p_2) where $f(p) \equiv 0$.*

PROOF:

If not, let (p_1, p_2) be the largest such gap and let (p^i) and (p^j) be sequences of prices in the support of f converging to p_1 and p_2, respectively. Then, $\lim F(p^i) = \lim F(p^j)$ since F is continuous, which implies

$$\frac{\pi_f(p_1)}{\pi_f(p_1) - \pi_s(p_1)} = \frac{\pi_f(p_2)}{\pi_f(p_2) - \pi_s(p_2)}$$

According to Proposition 5 this is impossible unless $p_1 = p_2$.

PROPOSITION 9: *If each store's optimal strategy involves zero probability of a tie, and $f(p) > 0$ for all $p^* \leqq p < r$, then each store must choose the same strategy.*

PROOF:

Let $F_i(p)$ be the optimal strategy for store $i = 1, \ldots, n$. Then by the reasoning of Proposition 4, stores k and j must satisfy the equations

---

[8] If all consumers are informed, only the lowest-price store will have any customers, and there can be only one price charged in equilibrium. But this is impossible by Proposition 2.

[9] More detailed empirical data on sales behavior of appliance retailers is presented in my working paper.

$$\pi_s(p)\prod_{i\neq j}(1-F_i(p))^{n-1}$$

$$=-\pi_f(p)\prod_{i\neq j}\left[1-(1-F_i(p))^{n-1}\right]$$

$$\pi_s(p)\prod_{i\neq k}(1-F_i(p))^{n-1}$$

$$=-\pi_f(p)\prod_{i\neq k}\left[1-(1-F_i(p))^{n-1}\right]$$

Dividing one equation into the other, we have

$$\frac{(1-F_k(p))^{n-1}}{(1-F_j(p))^{n-1}}=\frac{1-(1-F_k(p))^{n-1}}{1-(1-F_j(p))^{n-1}}$$

which implies $F_j(p)=F_k(p)$.

REFERENCES

G. Butters, "Equilibrium Distribution of Sales and Advertising Prices," *Rev. Econ. Stud.*, Oct. 1977, *44*, 465–91.

J. Pratt, D. Wise, and R. Zeckhauser, "Price Variations in Almost Competitive Markets," *Quart. J. Econ.*, May 1979, *93*, 189–211.

M. Rothschild, "Models of Markets with Imperfect Information: A Survey," *J. Polit. Econ.*, Dec. 1973, *81*, 1283–308.

S. Salop, "The Noisy Monopolist: Imperfect Information, Price Dispersion and Price Discrimination," *Rev. Econ. Stud.*, Oct. 1977, *44*, 393–406.

______ **and J. Stiglitz**, "A Theory of Sales," mimeo., Stanford Univ. 1976.

______ **and** ______, "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Rev. Econ. Stud.*, Oct. 1977, *44*, 493–510.

Y. Shilony, "Mixed Pricing in Oligopoly," *J. Econ. Theory*, Apr. 1977, *14*, 373–88.

G. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 213–25.

J. Stiglitz, "Equilibrium in Product Markets with Imperfect Information," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 339–45.

H. Varian, "A Model of Sales," work. paper no. C-11, Univ. Michigan 1978.

L. Wilde and A. Schwartz, "Equilibrium Comparison Shopping," *Rev. Econ. Stud.*, July 1979, *46*, 543–54.

An Alternative Test of the Capital Asset Pricing Model

By PAO L. CHENG AND ROBERT R. GRAUER*

The capital asset pricing model (*CAPM*) of William Sharpe (1964) and John Lintner (1965), later extended by Fischer Black (1972), has been extensively tested by Black, Michael Jensen and Myron Scholes, Marshall Blume and Irwin Friend, and Eugene Fama and James MacBeth.[1] The results of these efforts were recently examined and criticized by Richard Roll who astounded financial economists by declaring that: "(a) No correct and unambiguous test of the theory has appeared in the literature, and (b) there is practically no possibility that such a test can be accomplished in the future" (1977a, p. 129).

While we concur with Roll's indictment in (a), we wish to take exception to his prediction in (b). In this paper, we propose an alternative test of the *CAPM* and report our test results that are free from the ambiguity imbedded in the past tests. In Section I, we discuss the sources of ambiguity of the traditional tests. In Section II, an attempt is made to develop unambiguously testable implications of the *CAPM*. Section III contains the test design and data, and the results are reported in Section IV.

I. The Sources of Ambiguity in the Traditional *CAPM* Tests

The fundamental source of ambiguity in the *CAPM* tests to date can be traced to the test's preoccupation with the linear relation between risk and return,[2] and the resulting specification of the return-generating process. Inasmuch as the *CAPM* is a one-period model, there is no guarantee from the theory that the risk-return linearity is stationary through time. It merely asserts that, for a given period t in a mean-variance optimizing world, if the investors have homogeneous beliefs, where μ_t is the vector of mean returns at t, and Σ_t is the matrix of return variances and covariances, then the mean return of an asset or portfolio j is linear with respect to its systematic risk β_{jt}. This is, of course, the familiar security market line (*SML*):

$$\mu_{jt}=\bar{r}_{zt}+(\bar{r}_{mt}-\bar{r}_{zt})\beta_{jt}=\bar{r}_{zt}(1-\beta_{jt})+\bar{r}_{mt}\beta_{jt} \tag{1}$$

where $\bar{r}_{zt}$ may be taken as the riskless return in t if the *CAPM* of Sharpe and Lintner is entertained, but is the mean return of the zero *beta* portfolio in the extended *CAPM* of Black;[3] $\bar{r}_{mt}$ is the mean return of the market portfolio in t; and β_{jt} is the systematic risk of security j. It is worth noting that μ_t and Σ_t are exogenous while $\bar{r}_{zt}$, $\bar{r}_{mt}$, and β_{jt} are endogenous, and the *SML* in (1) is an equilibrium relationship, not a function.

Suppose (1) holds for each $t\in T$, where T is the sampling period. Consider now the return-generating process postulated for the purpose of statistical estimation in the past tests of the *CAPM*:[4]

---

*Department of economics and commerce, Simon Fraser University. We thank George Borts and the referee for helpful comments, and Doug McKinnon for programming assistance. The research was supported by a Canada Council grant.

[1] There are a number of empirical papers in existence other than those cited. See, for example, the references in Jensen. The cited papers, however, represent the most substantial and sophisticated testing of the *CAPM* to date.

[2] The preoccupation is understandable if one realizes that a risk and return relationship was a concern of the students of finance long before the *CAPM* made its appearance.

[3] It should be pointed out that in the Sharpe-Lintner *CAPM*, the riskless return is exogenous to the model, while the mean of Black's zero-*beta* portfolio is definitely endogenous. More of this will be dealt with in Section III.

$$(2)\qquad \tilde{r}_{jt}=\tilde{r}_{zt}(1-\beta_j)+\tilde{r}_{mt}\beta_j+\tilde{e}_{jt},\ t\in T$$

where the tildas denote random variables that follow a stationary stochastic process, and the mean of $\tilde{e}_{jt}$ and its covariance with other random variables are assumed to equal zero. Note that *beta* is no longer time subscripted, implying that β_j is a constant for all $t\in T$, i.e., a parameter to be estimated.[5] The price one pays for this move is to ignore the fact that β_j is endogenously determined for each t.

A fixed β_j for all $t\in T$ must also necessarily imply that the equilibrium firm values must vary through time in equal proportions. To see this, let us express (1) as

$$(3)\qquad \beta_{jt}=\frac{1}{\bar{r}_{mt}-\bar{r}_{zt}}\mu_{jt}-\frac{\bar{r}_{zt}}{\bar{r}_{mt}-\bar{r}_{zt}}$$

which makes β_{jt} a function of $\tilde{r}_{zt}$.[6] Thus, if $\beta_{jt}=\beta_j$, $t\in T$, then (2) and (3) together imply

$$(4)\qquad \beta_j=\frac{1}{\bar{r}_m-\bar{r}_z}\mu_j-\frac{\bar{r}_z}{\bar{r}_m-\bar{r}_z}$$

which is also seen as the expectation of (2). This in turn implies that $\bar{r}_m$, $\bar{r}_z$, and μ_j are fixed for all $t\in T$. In order for $\bar{r}_m=\bar{r}_{mt}=p_t'\mu/p_t'q$, where p_t' is the row vector of the equilibrium values of firms at t, μ is the column vector of mean returns fixed across time, and q is a column vectors of one's, p_t would have to satisfy the constraint $p_t=k_t\bar{p}$, where $\bar{p}$ is any column vector independent of t and k_t is a scalar depending on t. That is, all prices would have to move in equal proportions through time,[7] an unreasonable price to pay for entertaining a return-generating process as specified in (2).

If we take (2) at its face value where $\tilde{r}_{zt}$ is made to play the role of a random variable while β_j is fixed, then (2) represents at least a misspecification as well as a contradiction. It is a contradiction because β_{jt} is a function of $\tilde{r}_{zt}$ and therefore could not be held constant, and it is a misspecification because β_j, as a population regression coefficient, is not independent of the regressor $\tilde{r}_{mt}$.

Moreover, with the β_js fixed as the "true" *betas* over the sampling period, there would exist, by implication, a corresponding true and fixed market portfolio x_m^*. Let β^* be the vector of true *betas* so that $\beta^*=\Sigma x_m^*/x_m^{*\prime}\Sigma x_m^*$, with $\Sigma=\Sigma_t$, $t\in T$. But the *CAPM* does not guarantee the existence of β^*, x_m^* and Σ for all $t\in T$. Suppose the *CAPM* holds for each $t\in T$, with the corresponding β_t, x_{mt}, and Σ_t satisfying (1). Should these β^*, x_m^*, and Σ be interpreted as some weighted averages of β_t, x_{mt}, and Σ_t, respectively? Note that if $x_m^*\neq x_{mt}$, then one would have the same problem of ambiguity due to the use of a proxy market portfolio as forcefully discussed by Roll (1977a, pp. 136–40). Moreover, the statement by Roll that "The theory is not testable unless the exact composition of the true market portfolio is known and used in the tests. This implies that the theory is not testable unless all individual assets are included in the sample" (1977a, p. 130) does not get to the root source of ambiguity in the linearity tests of the *CAPM*.[8] For example, suppose that the *CAPM* holds for each t according to (1). The use of a portfolio of fixed weights for all $t\in T$ could result in the rejection of the *CAPM*, whether one calls such a portfolio the true market portfolio or a proxy, because the true market portfolio may not be the market portfolio for some or all t.

Finally, it is worth pointing out that the imposition of fixed β in the traditional *SML* tests carries with it the assumption of return distribution stationarity as well as the assumption of constant $\tilde{r}_{zt}$. With $\Sigma_t=\Sigma$ and

---

[4] See Black, Jensen, and Scholes (p. 99). The specifications by Fama and MacBeth (p. 611) and Blume and Friend (p. 21) are for all intents and purposes equivalent to (2).

[5] See Black, Jensen, and Scholes (p. 86) where the *betas* are assumed stationary, based on evidence presented by Blume.

[6] It is well known that $\tilde{r}_{mt}$ and $\tilde{r}_{zt}$ are orthogonal. Hence $\tilde{r}_{mt}$ is a function of $\tilde{r}_{zt}$ as well. See Roll (1977a).

[7] See Grauer for additional discussion on this point. Also see James Ohlson and Barr Rosenberg for a similar concern.

[8] Roll (1977b, p. 4) did recognize this point, but did not state the problem quite this way.

$\mu_t=\mu$, and defining $a=q\Sigma^{-1}\mu, b=\mu'\Sigma^{-1}\mu$, and $c=q'\Sigma^{-1}q$, we can write, upon substituting the relationship $\bar{r}_{mt}=(b-\bar{r}_{zt}a)/(a-\bar{r}_{zt}c)$ in (3),

$$\beta_t=\left(\frac{b-\bar{r}_{zt}a}{a-\bar{r}_{zt}c}-\bar{r}_{zt}\right)^{-1}(\mu-\bar{r}_{zt}q)$$

It is clear that in order for $\beta_t=\beta$, $\bar{r}_{zt}$ would have to be constant and independent of t. It is equally clear that the assumption of stationary return distributions does *not* imply stationary β_t. Since r_{zt}, being endogenous, is a function of the investors' initial wealth, their risk aversion, the size of the market, and the lending-borrowing opportunities,[9] stationary β represents an exceedingly strong assumption. Thus the past tests of the *CAPM* appear to have the unique methodological distinction of attempting to hold constant quantities that are not constant from equilibrium to equilibrium.

We therefore conclude that the postulated return-generating process in the *SML* tests contains a number of assumptions that imply a *CAPM* disconcertingly different from the *CAPM* being tested. Consequently, the test results would have to be ambiguous and the source of ambiguity would appear to be more fundamental than that which caused Roll's (1977a) consternation. Nevertheless, the *CAPM* needs to be tested. Fortunately, there exists a proposition that is 1) equivalent to the *SML*, and 2) capable of inducing propositions that can be unambiguously tested. Moreover, these propositions can be tested with a parsimony of assumptions.

II. An Alternative Test of the *CAPM*

A. *Theory*

In order to free ourselves from the preoccupation with risk-return linearity, we begin by rewriting the *SML* in (1) in accordance with the standard practice in economic analysis: that is, in equilibrium, endogenous variables are functions of exogenous parameters. The key endogenous variables in the *CAPM* are of course the values of firms. In vector form we rewrite (1) as follows:

$$\begin{aligned}\mu_t&=\bar{r}_{zt}q+(\bar{r}_{mt}-\bar{r}_{zt})\beta_t\\&=\bar{r}_{zt}q+\left(\frac{p_t'\mu_t-\bar{r}_{zt}p_t'q}{p_t'q}\right)\left(\frac{\Sigma_t p_t/p_t'q}{p_t'\Sigma_t p_t/(p_t'q)^2}\right)\\&=\bar{r}_{zt}q+\left(\frac{p_t'\mu_t-\bar{r}_{zt}p_t'q}{p_t'\Sigma_t p_t}\right)\Sigma_t p_t\\&=\bar{r}_{zt}q+\gamma_t\Sigma_t p_t\end{aligned}$$

where γ_t is the familiar "market price of risk."[10] Premultiplying by Σ_t^{-1} and rearranging terms yields

$$p_t=\gamma_t^{-1}(\Sigma_t^{-1}\mu_t-\bar{r}_{zt}\Sigma_t^{-1}q) \tag{5}$$

where both γ_t^{-1} and $\bar{r}_{zt}$ are some functions of exogenous variables.[11] Hence, as it should in equilibrium, (5) has been cast in the form desired but is equivalent to the *SML* in (1).

All the past tests of the *CAPM* have assumed stationarity in the joint distributions of security returns.[12] We shall make such an assumption and it is the only assumption we shall need in order for (5) to yield testable propositions. By setting $\Sigma_t=\Sigma$ and $\mu_t=\mu$ for all $t\in T$, T being the sampling period, we have that

$$p_t=\gamma_t^{-1}(\Sigma^{-1}\mu-\bar{r}_{zt}\Sigma^{-1}q),\ t\in T \tag{6}$$

where $\Sigma^{-1}\mu$ and $\Sigma^{-1}q$ are seen as vectors independent of t. Thus, if the joint distributions of security returns are stationary during a sampling period T, then (6) must hold for all $t \in T$. Without additional assumptions, we shall show below that (6) can be unambiguously tested, due at least in part to a lack of the need to estimate $\gamma_t, \bar{r}_{zt}, \Sigma$, or μ.

---

[9] This will be further discussed in Section II, but see also Cheng (1977; forthcoming).

[10] See, for example, Jan Mossin and Lintner (1971) for the determination of market price of risk.

[11] See Cheng (1977; 1978; forthcoming) and John Heaney for a detailed discussion of γ_t^{-1} and $\bar{r}_{zt}$ as some explicit or unspecified funtions of exogenous variables.

[12] We are in sympathy with Fama (1977) who states: "...without a large degree of stationarity in the joint distribution of security returns, empirical testing of the SLB [Sharpe-Lintner-Black] model would likely be ruled out" (p. 22). However, as we have shown in Section I, if the testing of the *CAPM* follows the *SML* route, other assumptions would have to be made in addition to the stationarity of the joint distribution of security returns.

Define $u=\Sigma^{-1}\mu$ and $v=\Sigma^{-1}q$ and write (6) as

$$(6') \qquad p_t=\gamma_t^{-1}(u-\bar{r}_{zt}v) \quad t \in T$$

where p_t is seen as a linear combination of time-independent vectors u and v. Consider the kth and the jth equation of (6′):

$$(7) \qquad p_{kt}=\gamma_t^{-1}(u_k-\bar{r}_{zt}v_k)$$

$$(8) \qquad p_{jt}=\gamma_t^{-1}(u_j-\bar{r}_{zt}v_j)$$

Solve (7) for $\bar{r}_{zt}$ and substitute it into (8) to obtain

$$(9) \quad p_{jt}=\gamma_t^{-1}\left(u_j-\frac{u_k v_j}{v_k}\right)+\frac{v_j}{v_k}p_{kt} \qquad j \neq k$$

where p_{jt} is linear with respect to p_{kt}. Now select another firm i, $i \neq j$, and write

$$(10) \quad p_{it}=\gamma_t^{-1}\left(u_i-\frac{u_k v_i}{v_k}\right)+\frac{v_i}{v_k}p_{kt} \quad i \neq k$$

Using (9) and (10) to eliminate the unobservable γ_t^{-1} and rearranging yields

$$(11) \quad p_{it}=\left(\frac{u_i v_k-v_i u_k}{u_j v_k-v_j u_k}\right)p_{jt} + \left[\frac{v_i}{v_k}-\frac{v_j}{v_k}\left(\frac{u_i v_k-v_i u_k}{u_j v_k-v_j u_k}\right)\right]p_{it} \equiv b_{ijk}p_{jt}+c_{ijk}p_{kt}$$

which makes the value of firm i linear homogeneous with respect to any two other arbitrary firm values. Since the coefficients are functions of the elements in Σ and μ, then any triplet of firm values observed at any $t \in T$ must have a fixed structure as in (11). That is, although p_{it}, p_{jt}, and p_{kt} may vary during T, the *CAPM* with stationary beliefs must necessarily imply that they should vary in lockstep. It is for this reason that (11) has been referred to by Cheng (forthcoming) as the Invariance Law of Relative Prices, and otherwise referred to as the Random March,[13] in deference to the traditional random walk hypothesis of stock price movements.

Several extensions of (11) can be made in anticipation of the test design to be proposed in Section III.

(a) The Invariance Law can be extended to relate more than three firm values. Consider the values of two additional firms, g and h, and their invariant relationship with the value of firm j:

$$(12) \quad p_{jt}=\left(\frac{u_j v_h-v_j u_h}{u_g v_h-v_g u_h}\right)p_{gt} + \left[\frac{v_j}{v_h}-\frac{v_g}{v_h}\left(\frac{u_j v_h-v_j u_h}{u_g v_h-v_g u_h}\right)\right]p_{ht} \equiv b_{jgh}p_{gt}+c_{jgh}p_{ht}$$

Substituting for p_{jt} in (11), from (12) we have that

$$(13) \quad p_{it}=(b_{ijk}b_{jgh})p_{gt}+(b_{ijk}c_{jgh})p_{ht}+c_{ijk}p_{kt}$$

which is again a linear homogeneous equation. Obviously, this regenerative scheme can be repeatedly applied to include additional values of firms to the extent desired.

(b) The Invariance Law can be extended to deal with a portfolio of firms. Using (6′), we choose three groups of firm values from p_t. Let n_1, n_2, and n_3 be the number of firms in the three groups. Then group J, $J=1,2,3$, will have a portfolio value

$$(14) \quad p_{Jt}=\sum_{j=1}^{n_J} p_{jt}=\gamma_t^{-1}\left(\sum_{j=1}^{n_J} u_j-\bar{r}_{zt}\sum_{j=1}^{n_J} v_j\right) \equiv \gamma_t^{-1}(u_J-\bar{r}_{zt}v_J) \qquad J=1,2,3$$

It is therefore obvious that (14) will yield us a linear homogeneous relationship among three portfolio values of firms similar to (11).

---

[13] The catchy description is due to Heaney, who in turn also generalized the Invariance Law. A test of the *CAPM* based upon the generalized Invariance Law will be reported in a forthcoming paper.

(c) The Invariance Law can be extended to relate security prices rather than firm values. Suppose during a sampling period T the change in the number of shares outstanding of firm j, n_j, is negligible for all j. Denote by N^{-1} the inverse of a diagonal matrix with n_js along the diagonal so that $N^{-1}p_t$ is the vector of security prices. It is obvious that the vector of security prices will be a linear combination of the time-independent vectors $N^{-1}u$ and $N^{-1}v$. Hence the Invariance Law must hold also for the prices of securities.[14]

B. *Testable Implications*

In view of our preceding discussion, we can state in general that the *CAPM*, under stationary return distributions, implies a linear homogeneous structure of equilibrium security prices, or firm values, or share values of firms. For a collection of $k+1$ firm values at t, the structure has the form:[15]

$$p_{it}=b_1p_{1t}+\ldots+b_kp_{kt},\ k\geqslant 2,\ \mathrm{t}\in T \tag{15}$$

where p_{it} is the $(k+1)$th firm value. Note that the $k+1$ different firm values can be arbitrarily chosen from the population of firms in T, and that $b_1,\ldots,b_k$ are constants, but are some functions of the components of the vectors $\Sigma^{-1}\mu$ and $\Sigma^{-1}q$. If an equilibrium exists at each $\mathrm{t}\in T$, then it does not matter whether the firm values are observed daily, weekly, or monthly. Consider the stochastic generalization of (15) of the form

$$p_{it}=b_{0k}+b_1p_{1t}+\ldots+b_kp_{kt}+e_{it},\ \mathrm{t}\in T \tag{16}$$

where e_{it} is the usual error term in a least squares regression and b_{0k} is an intercept term subscripted by the number of independent variables to indicate that its estimated value could conceivably vary depending on the value of k. We focus on five testable implications of the model in this paper.

In terms of the regression equation (16), with $k=2$, the Invariance Law implies that the following must hold *jointly*:

HYPOTHESIS 1: The intercept equals zero: $b_{0k}=b_{02}=0$

HYPOTHESIS 2: The slope coefficients differ from zero: $b_1, b_2\neq 0$.[16]

---

[14] There exists, however, the reasonable concern that the values of firms should include the values of bonds. Fortunately, the law still holds when corporate bonds are included as risky securities. For the sake of discussion, suppose there exists a single class of bonds for each of the N firms. Since there will be $2N$ risky securities, Σ will have a dimension of $2N$ by $2N$ and μ a dimension of $2N$ by one. The valuation equation (6), i.e., $p_t=\gamma_t^{-1}(\Sigma^{-1}\mu-\bar{r}_{zt}\Sigma^{-1}q)$, can be interpreted accordingly. That is, for example, p_t will be a $2N$-by-one vector containing N values of shares and N values of bonds. Hence, the Invariance Law should hold separately for the share values of firms as well as the bond values of firms. Indeed, in the empirical work to follow in Section III, we shall only use the share values of firms for testing the *CAPM*.

[15] See fn. 14. We shall hereafter use the term "firm values" to mean share values of firms. Moreover, equation (15) as a hypothesis to be tested is much less restrictive than the two-factor *SML* in (1). As pointed out by the referee of this paper, (6) remains a linear combination of two time-independent vectors if we generalize it by writing

$$p_t=\theta_t\gamma_t^{-1}[\Sigma^{-1}(\mu+\phi)-\bar{r}_{zt}\Sigma^{-1}\psi]$$

where θ_t is an arbitrary nonzero scalar, and ϕ and ψ are arbitrary vectors of constants ($\psi\neq 0$). The above implies the hypothesis,

$$\mu=-\phi+\bar{r}_{zt}\psi+\theta_t(\bar{r}_{mt}-\bar{r}_{zt})\beta$$

Of course, if this hypothesis, or equivalently (15), is rejected, then so is the *SML* in (1). Although difficult to implement, a more powerful test would require restrictions on the coefficients in (15).

[16] To see that the coefficient cannot be zero, refer to (11) for the expression of $b_1=b_{ijk}$ and $b_2=c_{ijk}$. Consider first $b_1=b_{ijk}$, where neither the numerator nor the denominator equal zero. For instance, if the numerator were equal to zero, then $u_i/u_k=v_i/v_k$. But u_i and u_k are the respective ith and kth component of $\Sigma^{-1}\mu$ while v_i and v_k are the respective ith and kth component of $\Sigma^{-1}q$. In order for $u_i/u_k=v_i/v_k$, all the components in μ would have to be identical, which is hardly a case of concern. To show that $b_2=c_{ijk}\neq 0$, note first:

$$b_2=\frac{v_i}{v_k}-\frac{v_j}{v_k}b_1$$

If $b_2=0$, then $v_i/v_j=b_1$. But b_1 can be written

$$\frac{v_i[(u_i/v_i)v_k-u_k]}{v_j[(u_j/v_j)v_k-u_k]}$$

HYPOTHESIS 3: The adjusted coefficient of determination, $R_k^2 = R_2^2$, should be near one.

As we add regressors to (16), two additional hypotheses can be identified:

HYPOTHESIS 4: No trend in the intercept: As regressors are added to (16) the intercepts for $k \geqslant 2$ should be equal to zero and should not deviate from zero as k increases from 2.

HYPOTHESIS 5: No trend in the adjusted coefficient of determination: As regressors are added to (16) the adjusted coefficient of determination of $k \geqslant 2$ should be near one and should remain near one as k increases from 2.

Note, however, that Hypotheses 1–5 are not independent. If one is rejected, then so is *CAPM*.

To operationalize Hypotheses 4 and 5 we first run regressions of the form given by (16), letting the number of independent variables increase from two to nineteen. Then, we take the intercepts and adjusted coefficients of determination from (16) and run the following regressions:

$$b_{0k} = \alpha_0 + \alpha_1 k + e \qquad k = 2, \ldots, 19 \tag{17}$$

$$R_k^2 = \beta_0 + \beta_1 k + e \qquad k = 2, \ldots, 19 \tag{18}$$

The testable implications of Hypothesis 4 are that both α_0 and α_1 equal zero and the testable implications of Hypothesis 5 is that β_1 equals zero.

III. The Test Design and the Data

The data for this study are monthly values of equity (the monthly closing price times the number of shares outstanding) for firms traded on the New York Stock Exchange during the period January 1926 through December 1977. The data are from the Center for Research in Security Prices (*CRSP*) of the University of Chicago.

Either individual security or portfolio data could be employed as input for the tests. In fact, one of the advantages of the tests is that they can be carried out on any subset of the data. But the three most widely quoted studies of the linear risk-return tradeoff predicted by the *CAPM* (Black, Jensen, and Scholes; Blume and Friend; Fama and MacBeth) have employed portfolio data for a number of econometric reasons.[17] While this paper is not confronted with the same econometric difficulties, we purposely employed a portfolio grouping technique so that the test results can be more easily compared with the previous tests of the *CAPM.* The actual grouping technique employed here corresponds most closely to the Fama-MacBeth approach. They used three periods to construct a time-series of risk-return data needed to test for the *CAPM*'s predicted linear risk-return tradeoff: a formation period (to group securities into portfolios based on individual security *betas*), an estimation period (to reestimate the individual security *betas* used in calculating the portfolio *betas*), and a test period (to calculate the portfolio returns and to actually perform the tests).

In detail, the following steps were undertaken in forming the monthly portfolio value data:

1) In each year, for which there were at least thirty-six observations from the previous sixty months (the formation period), and twelve observations from the twelve-month test period that followed the intervening sixty-month estimation period, the

---

to show that if $v_i/v_j = b_1$, then u_i/v_i must equal u_j/v_j. But this is possible only if all the components in μ are identical. For the purpose of this study the specific values of b_1 and b_2 are not of concern, only that they are significantly different from zero.

[17] The primary reason these studies employed the grouping procedure was to avoid an errors-in-the-variables, or measurement error problem, in the *betas* used in the second pass regressions testing for the linearity, slope, and intercept of the security market line. However, the grouping technique was also used because it was felt that it may give more efficient *ex post* estimates of *ex ante* conditional return distributions and it provides an easy way to deal with delistings.

betas were calculated over the formation period.[18] (The first formation period was forty-eight months in length so that the portfolio data could be generated from January 1935, the beginning month of the Fama-MacBeth study.)

2) The stocks were then ranked from largest to smallest based on the *betas* calculated in the formation period.

3) Twenty portfolios were selected each year by assigning the highest 5 percent of the ranked stocks to portfolio 1, the next 5 percent to portfolio 2, and so on. (More accurately, if n stocks were available for selection in year t, $n/20$ stocks were assigned to each of the twenty portfolios. The remainder from $j = \text{mod}\ (n, 20)$ was distributed such that one extra stock was included in each of the first j portfolios.)

4) Over the twelve-month test period, the portfolio values were calculated as the sum of the prices times the number of shares outstanding of the individual securities contained in the portfolios in each month.

5) Steps 1 to 4 were repeated for each year from January 1930 until January 1972 so that the sets of monthly values for the twenty portfolios were created for the forty-three-year (516-month) period beginning in January 1935 and ending in December 1977.

In designing the tests we endeavored to provide a wide range of dependent and independent variables for the regressions (particularly for the regressions containing only two independent variables). Portfolios 1, 10, and 20 were chosen as dependent variables. Portfolio 10 was chosen as it had a β value near 1. On the other hand, portfolios 1 and 20 were chosen as they provided the opposite extremes: portfolios containing the lowest and highest β securities, respectively. Having picked one of three dependent variables in any time period, regressions containing two independent variables were run for nine different sets of values of the independent variables. For the values of firms in portfolios 10 and 20 as the dependent variables, the independent variables were the values of firms in portfolios j and k, $j = 1, \ldots, 9$ and $k = 20 - j$. For the value of firms in portfolio 1 as the dependent variables, the independent variables were the values of firms in portfolios j and k, $j = 2, \ldots, 10$ and $k = 21 - j$. The j, k combinations were also chosen to provide a wide variety of possible β portfolios. For example, with portfolio 10 as the dependent variable the regressions range from the high and low β portfolios 1 and 19 as the independent variables to the contiguous β portfolios 9 and 11.

IV. Results

The regression results based on equation (16) for two regressors over periods of various lengths are presented in Table 1. The various periods are: the full 516-month period, four 129-month nonoverlapping subperiods, and the 402-month Fama-MacBeth period. For each period, nine regressions were run on each of the three selected regressants: portfolios 1, 10, and 20. For each period and each regressant, Table 1 gives the summary information on the associated set of nine regressions: the averages of the absolute values of nine t-statistics, denoted by $|\bar{t}(\hat{b}_0)|$, $|\bar{t}(\hat{b}_1)|$, and $|\bar{t}(\hat{b}_2)|$, respectively, for the intercept $\hat{b}_0$ and the two slope coefficients $\hat{b}_1$ and $\hat{b}_2$, and the average adjusted coefficient of determination, denoted by $\bar{R}^2$.

It is not clear whether the Invariance Law is more likely to hold over short or long periods. Some argue that beliefs are likely to remain stationary only over short periods. Yet others hold a somewhat different view

---

[18] The primary difference between the grouping technique employed by the *SML* studies and the technique employed here is our requirement that for a security to be included in the study there must be a full data complement for the test period. The *SML* authors argued that the exclusion of securities delisted during the test period might introduce a survivorship bias. It was simple for them to deal with the problem by constructing the time-series of portfolio *betas* and returns as averages of the *betas* and returns of the individual securities contained in a portfolio during a particular month. While we appreciate the survivorship argument, the reader will recall that it is not necessary to include all securities in our study—the Invariance Law can be tested on any subset of data. In any event, because our goal is to construct portfolio values, it is not possible to employ the averaging technique used in the *SML* studies. For those particularly interested in how closely the data sets interrelate, in the 1960's the Fama-MacBeth data sets contained on the order of 850 securities while our data sets contained on the order of 650 securities.

TABLE 1—SUMMARY STATISTICS FROM THE REGRESSIONS[a]

$p_{it}=b_0+b_1p_{jt}+b_2p_{kt}+e_{it}$

Time Period	Regressant Portfolio	$\lvert\bar{t}(\hat{b}_0)\rvert$	No. of Times $\lvert\bar{t}\rvert>2$	$\lvert\bar{t}(\hat{b}_1)\rvert$	No. of Times $\lvert\bar{t}\rvert<2$	$\lvert\bar{t}(\hat{b}_2)\rvert$	No. of Times $\lvert\bar{t}\rvert<2$	$\bar{R}^2$
1/35–9/45	1	9.28	9	4.50	2	3.94	4	0.26
	10	5.42	7	4.43	1	4.09	5	0.23
	20	6.95	8	2.88	4	4.19	2	0.18
10/45–6/56	1	5.27	7	13.31	0	5.48	1	0.74
	10	6.30	9	10.62	1	2.82	4	0.65
	20	7.55	9	8.91	1	4.75	2	0.59
7/56–3/67	1	4.43	7	9.76	0	3.73	2	0.48
	10	2.65	5	6.51	0	4.89	2	0.41
	20	9.50	9	4.97	1	6.00	2	0.28
4/67–12/77	1	11.59	9	6.81	0	3.61	5	0.34
	10	11.07	9	5.82	2	4.02	3	0.33
	20	7.55	9	5.58	2	4.12	3	0.32
1/35–6/68	1	3.75	8	28.08	0	10.22	1	0.83
	10	2.49	6	20.05	0	8.45	1	0.72
	20	10.85	9	14.78	0	4.88	2	0.45
1/35–12/77	1	6.91	9	26.29	0	14.04	1	0.79
	10	3.34	6	19.06	0	9.63	0	0.67
	20	5.87	9	5.86	3	12.33	1	0.45

Note: Additional and detailed statistical results from the regressions will be supplied by the authors upon request.

[a] Based on *beta* ranked portfolios for selected periods during January 1935–December 1977.

TABLE 2—SUMMARY AND AVERAGE STATISTICS FROM THE REGRESSIONS[a,b]

Average	Average $\lvert\bar{t}(\hat{b}_0)\rvert$	Avg. No. of Times $\lvert\bar{t}\rvert>2$	Average $\lvert\bar{t}(\hat{b}_1)\rvert$	Avg. No. of Times $\lvert\bar{t}\rvert<2$	Average $\lvert\bar{t}(\hat{b}_2\rvert$	Avg. No. of Times $\lvert\bar{t}\rvert<2$	Avg. $\bar{R}^2$
Five-year Period:	8.01	7.63	5.00	2.88	5.38	2.58	.43
Three-year Period:	7.51	6.98	8.17	1.74	7.91	1.62	.71

Note: See Table 1.

[a] The averages are computed over twenty-four five-year periods and forty-two three-year periods.

[b] Based on *beta* ranked portfolios for the five-year and three-year periods during January 1935–December 1977.

that beliefs are only likely to be realized over longer periods. As both arguments have a considerable number of adherents, as well as certain plausibility, we also ran regressions over shorter periods: five- and three-year nonoverlapping periods during January 1935–December 1977. Using portfolios 1, 10, and 20 as regressants for each period, a total of twenty-four sets of regressions for the five-year periods and a total of forty-two sets for the three-year periods (with nine regressions in each set) were run. Since it is cumbersome to report the results in full, Table 2 summarizes the average characteristics of the two sets of short-period regressions. That is, the table reports statistics over the twenty-four sets of five-year regressions and over the forty-two sets of the three-year regressions.

Consider Hypothesis 1: the intercept should not differ from zero. A glance at Tables 1 and 2 shows that the averages of the absolute values of the t-statistics are much larger than 2. In 89 percent of the regressions in Table 1, 85 percent of the five-year regressions and 78 percent of the three-year regressions in Table 2, or in 82 percent of the 756 regressions performed overall, the intercepts had t-values in excess of 2. Turn to Hypothesis 2: the slope coefficients should differ from zero. We find that overall 21 percent of the $\hat{b}_1$ and 23 percent

TABLE 3—A COMPARISON OF ORDINARY LEAST SQUARES (*OLS*) AND COCHRANE-ORCUTT (C-O) REGRESSIONS FOR THE 1935–77 PERIOD AND THE 1935–39 PERIOD[a]

	Regressor Portfolios	$t(\hat{b}_0)$	$t(\hat{b}_1)$	$t(\hat{b}_2)$	R^2 (unadjusted)	*D.W.* Statistic	ρ^b
1/35–12/77							
OLS	1 19	−.24	18.07	8.67	.78	.13	
C-O	1 19	2.26	8.16	3.31	.98	1.97	.97
OLS	2 18	2.11	27.21	3.52	.71	.12	
C-O	2 18	2.38	5.20	4.01	.97	1.98	.98
OLS	3 17	4.82	21.15	2.27	.62	.15	
C-O	3 17	2.70	−4.67	−3.33	.97	1.94	.99
OLS	4 16	4.40	23.32	3.98	.71	.11	
C-O	4 16	2.36	9.63	3.39	.98	1.98	.98
OLS	5 15	7.28	16.30	6.08	.49	.11	
C-O	5 15	2.40	.58	−7.45	.97	1.94	.99
OLS	6 14	4.22	14.08	18.98	.62	.10	
C-O	6 14	2.55	3.51	1.58	.97	1.96	.98
OLS	7 13	5.29	16.51	10.72	.59	.14	
C-O	7 13	2.58	−1.30	−2.75	.97	1.96	.99
OLS	2 12	1.62	19.46	14.75	.72	.19	
C-O	2 12	2.24	−3.90	1.56	.97	1.95	.99
OLS	9 11	−.06	15.44	17.68	.83	.24	
C-O	9 11	2.67	−2.24	.39	.97	1.96	.99
1/35–9/45							
OLS	1 19	7.77	9.41	−11.62	.61	.26	
C-O	1 19	6.15	7.68	−10.50	.91	1.94	.87
OLS	2 18	1.80	5.63	9.91	.61	.19	
C-O	2 18	−.61	6.27	12.49	.94	2.07	.94
OLS	3 17	7.63	2.23	.14	.05	.24	
C-O	3 17	4.23	−2.68	2.33	.82	2.01	.92
OLS	4 16	8.67	3.24	−1.85	.11	.23	
C-O	4 16	4.10	1.05	−.49	.80	1.96	.89
OLS	5 15	2.91	3.46	4.64	.19	.25	
C-O	5 15	2.66	2.86	1.67	.82	1.95	.88
OLS	6 14	4.61	2.24	1.36	.04	.21	
C-O	6 14	.99	1.31	5.46	.84	2.01	.94
OLS	7 13	−.44	7.63	4.72	.34	.25	
C-O	7 13	−.45	5.49	5.16	.86	1.93	.91
OLS	8 12	8.62	−.82	−1.47	.02	.19	
C-O	8 12	6.23	−3.88	−5.10	.84	2.08.94	
OLS	9 11	6.37	5.22	−1.08	.21	.28	
C-O	9 11	3.38	−.66	2.54	.82	1.97	.91

Note: See Table 1.

[a] The portfolio 10 is the regressant in each regression.

[b] The C-O ρ, the first-order serial coefficient.

of the $\hat{b}_2$ coefficients did not statistically differ from zero. Now consider Hypothesis 3: the adjusted coefficients of determination should be near one. The data show that the average adjusted coefficients of determination were more on the order of .5. In sum, the results can hardly be said to support the joint Hypotheses 1,2, and 3.

However, a word of caution is in order. All the regressions exhibited a high degree of serial correlation. Overall the Durbin-Watson statistics averaged on the order of .3 or .4. But the serial correlation was more acute when measured over longer periods, where the Durbin-Watson statistics were more on the order of .1. Serial correlation implies that the ordinary least squares (*OLS*) estimators of the regression coefficients are unbiased and consistent, but they are not efficient or asympotically efficient.

TABLE 4—SUMMARY RESULTS FOR THE REGRESSIONS
$b_{0k}=\alpha_0+\alpha_1 k+e, k=2,\ldots,19$; $\bar{R}^2_k=\beta_0+\beta_1 k+e, k=2,\ldots,19$

Time Period	Based on Dependent Variable Portfolio	α_0	$t(\hat{\alpha}_0)$	α_1	$t(\hat{\alpha}_1)$	$\bar{R}^2$	$\hat{\beta}_0$	$t(\hat{\beta}_0)$	$\hat{\beta}_1$	$t(\hat{\beta}_1)$	$\bar{R}^2$
1/35–9/45	1	127,970	4.71	−12,024	−5.19	.60	.54	14.16	.027	8.11	.79
	10	541,342	5.93	−8,739	−1.12	.02	.33	4.58	.044	7.19	.75
	20	2,257,686	2.82	−75,960	−1.11	.01	.21	2.22	.052	6.43	.70
10/45–6/56	1	−117,168	−4.88	15,588	7.61	.77	.92	66.57	.005	4.44	.52
	10	415,127	4.87	−9,830	−1.35	.05	.84	61.89	.009	8.18	.80
	20	1,496,831	2.83	80,020	1.77	.11	.58	19.86	.025	10.08	.86
7/56–3/67	1	699,069	4.18	−119,966	−8.40	.80	.60	24.72	.024	11.66	.89
	10	$-.104\times10^8$	−4.82	442,403	2.40	.22	.52	12.75	.031	8.77	.82
	20	$.137\times10^8$	10.89	31,208	.29	0.	.45	12.55	.033	10.69	.87
4/67–12/77	1	3,271,176	10.71	−249,039	−9.55	.84	.59	21.21	.023	9.58	.84
	10	$.250\times10^8$	9.84	−14,414,461	−6.52	.71	.60	10.47	.026	5.32	.62
	20	$.109\times10^9$	9.74	−6,364,723	−6.66	.72	.63	15.16	.024	6.58	.71
1/55–6/68	1	210,678	6.03	−15,604	−5.23	.61	.90	247.05	.005	16.34	.94
	10	−1,076,703	−5.47	−59,692	−3.55	.41	.80	53.32	.012	9.76	.85
	20	5,819,474	10.53	−244,661	5.18	.60	.51	24.82	.025	14.07	.92
1/35–12/77	1	311,371	13.29	−20,271	−10.13	.86	.88	209.22	.006	12.76	.95
	10	977,743	3.44	−156,804	−6.44	.70	.73	52.80	.014	12.07	.89
	20	8,715,239	25.80	−319,426	−11.07	.88	.14	1.87	.049	7.57	.77

Under the assumption that a first-order autoregressive scheme more accurately characterizes the data, the observation of highly positive serial correlation implies that the estimates of the variances of the least squares estimators are too small. This, in turn, implies that the *t*-statistics are overstated. In our case, taking the *t*-statistics at face value will lead us, when concentrating on the intercepts, to be more apt to reject the *CAPM* than we should. On the other hand, when we concentrate on the slopes, we will be more apt to accept the model when we should not. But the impressions gained from the *OLS* *t*-statistics may not be as biased as one might first expect. We ran a subset of the regressions again, using the Cochrane-Orcutt (C-O) iterative method to adjust for autocorrelation. The results of the C-O regressions, run over two periods, and the corresponding *OLS* regressions are shown in Table 3. Overall the results indicate that the mean absolute values of the *t*-statistics are smaller for the C-O regressions. However, the absolute values of the *t*-statistics for the intercept exceeded 2 fourteen times for the *OLS* regressions, and fifteen times for the C-O regressions. Turning to the slope coefficients, the absolute values of the *t*-statistic on $\hat{b}_1$ are less than 2, once for the *OLS* and five times for the C-O regressions. For $\hat{b}_2$ the absolute values of the *t*s are less than 2, four times for the *OLS* and five times for the C-O regressions. Thus, the *OLS* and C-O results lead to very similar results with respect to Hypotheses 1 and 2. On the other hand, the C-O regressions yield substantially higher (unadjusted) coefficients of determination. Considering Hypotheses 1, 2, and 3 as joint hypotheses, we see no reason to alter our conclusions reached on the basis of the *OLS* results.

The second set of hypotheses, 4 and 5, are concerned with whether there are any systematic changes in either the intercept or the adjusted coefficient of determination as the number of independent variables in (16) is increased. We proposed to test the proposition that there should be no systematic effect on either the intercept, or adjusted coefficient of determination, by regressing these variables against k, the number of regressors. The results of these two regressions, (17) and (18), covering various subperiods and the three reference portfolios 1, 10, and 20, are presented in Table 4.

TABLE 5—SUMMARY STATISTICS FROM THE REGRESSIONS[a]

Time Periods	Regressant Portfolio	$\lvert\bar{t}(\hat{b}_0)\rvert$	No. of Times $\lvert\bar{t}\rvert>2$	$\lvert\bar{t}(\hat{b}_1)\rvert$	No. of Times $\lvert\bar{t}\rvert<2$	$\lvert\bar{t}(\hat{b}_2)\rvert$	No. of Times $\lvert\bar{t}\rvert<2$	$\bar{R}^2$
1/35–9/45	1	3.26	6	7.92	5	5.58	1	0.76
	10	2.14	4	8.11	2	6.14	2	0.77
	20	4.73	6	9.29	1	4.05	6	0.77
10/45–6/56	1	6.42	7	5.41	2	9.77	0	0.97
	10	5.67	6	7.63	0	11.29	1	0.98
	20	4.15	7	6.20	1	14.80	0	0.98
7/56–3/67	1	12.36	8	6.71	2	5.66	1	0.84
	10	7.29	8	7.29	0	6.94	1	0.87
	20	13.07	8	5.78	2	4.10	2	0.71
4/67–12/77	1	5.11	9	6.26	2	5.96	2	0.38
	10	10.04	9	3.19	3	3.88	3	0.22
	20	4.57	7	8.21	0	3.79	3	0.47
1/35–6/68	1	18.49	9	18.25	1	16.66	0	0.96
	10	6.52	7	18.54	0	17.13	1	0.98
	20	15.95	9	17.17	1	14.43	2	0.95
1/35–12/77	1	17.41	9	19.88	1	13.34	0	0.87
	10	6.16	7	16.68	1	14.70	2	0.93
	20	20.50	9	19.79	1	11.89	1	0.90

Note: See Table 1.

[a] Based on fixed group portfolios for selected periods during January 1935–December 1977.

The results show that there are predominantly statistically significant trends in the estimated values of the intercept as regressors are added. However, the trends are not of the same sign. The results concerning trends in the adjusted coefficients of determination are clearer. There is a statistically significant increase in the adjusted coefficient of determination as the number of regressors increases. This appears to contradict Hypothesis 5 and the mean-variance model.[19]

The results we have reported up to this point are based on a grouping procedure that observes the *beta*-ranks of the portfolios of firms in order to provide comparability with the *SML* tests. However, as a check on whether our results are sensitive to that particular grouping procedure, we entertain an alternative grouping procedure that requires invariance through time in the group membership, without regard to the firms' *beta*. Thus we first harvested 253 firms from the *CRSP* tape that survived the entire January 1935–December 1977 period. We then selected the first 240 firms and organized twenty portfolios of 12 firms each, without regard to their *beta* ranks. The group membership of a firm remains unchanged throughout the forty-three years. In a format similar to Table 1, the regression results for the various periods are reported in Table 5. With the exception of $\bar{R}^2$, Tables 1 and 5 appear to give strikingly similar results. We also ran the similar regressions for the five-year and three-year periods and found the average results very similar to the results reported in Table 2.

V. Concluding Remarks

We have provided evidence against the *CAPM*. The five hypotheses we tested are generated from the Invariance Law, a proposition equivalent to, although perhaps less intuitively pleasing than, the *SML* proposition that characterizes a positive risk-return tradeoff. The primary virtue of the Invariance Law tests is that given return stationarity, the resulting evidence is unambiguous, which, as Roll (1977a) pointed out, cannot be claimed for the *SML* tests. The Invariance Law tests we have performed entertained nothing beyond, indeed substantially less than, what the *SML* tests have assumed. To the extent that the Invariance Law requires return stationarity, the *SML* tests require the same assumption. However, neither framework can cope with the possibility that the *CAPM* may hold for each period while the return distributions are nonstationary.

---

[19] This interpretation is in recognition of the fact that the distribution of the adjusted coefficient of determination is skewed and the degree of skewness changes with the number of regressors.

REFERENCES

F. Black, "Capital Market Equilibrium with Restricted Borrowing," *J. Bus., Univ. Chicago,* July 1972, *45,* 444–45.

______, **M. Jensen, and M. Scholes,** "The Capital Asset Pricing Model: Some Empirical Tests," in Michael Jensen, ed., *Studies in the Theory of Capital Markets,* New York 1972.

M. Blume, "The Assessment of Portfolio Performance," unpublished doctoral dissertation, Univ. Chicago 1968.

______ **and I. Friend,** "A New Look at the Capital Asset Pricing Model," *J. Finance,* Mar. 1973, *28,* 19–34.

P. Cheng, "Divergent Rates, Financial Restriction, and Relative Prices in Capital Market Equilibrium," *J. Finan. Quant. Anal.,* forthcoming.

______, "Capital Market Equilibrium with Endogenous Borrowing Rates," dept. econ. commerce disc. paper series no. 77-12.3, Simon Fraser Univ. 1977.

______, "Default Risks, Portfolio Debts and Corporate Bond Valuation," unpublished manuscript, 1978.

E. Fama, "Risk-adjusted Discount Rates and Capital Budgeting Under Uncertainty," *J. Finan. Econ.,* Aug. 1977, *5,* 3–24.

______ **and J. MacBeth,** "Risk, Return and Equilibrium: Empirical Tests," *J. Polit. Econ.,* May/June 1973, *81,* 607–36.

R. Grauer, "Belief Reinforcement in Capital Asset Pricing, with Implications for Empirical Testing," dept. econ. commerce disc. paper series no. 78-6-1, Simon Fraser Univ. 1978.

J. Heaney, "Prices of Risky Assets in General Equilibrium," unpublished masters thesis, Simon Fraser Univ. 1977.

M. Jensen, "Capital Markets: Theory and Evidence," *Bell J. Econ.,* Autumn 1972, *3,* 357–98.

J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Econ. Statist.,* Feb. 1965, *47,* 13–47.

______, "The Market Price of Risk, Size of Market and Investor's Risk Aversion," *Rev. Econ. Statist.,* Feb. 1971, *53,* 87–99.

R. Merton, "An Analytical Derivation of the Efficient Portfolio Frontier," *J. Finan. Quant. Anal.,* Sept. 1972, *7,* 1851–72.

J. Mossin, "Security Pricing and Investment Criteria in Competitive Markets," *Amer. Econ. Rev.,* Dec. 1969, *59,* 749–56.

J. Ohlson and B. Rosenberg, "The Stationary Distribution of Returns and Portfolio Separation in Capital Markets: A Fundamental Contradiction," *J. Finan. Quant. Anal.,* Sept. 1976, *11,* 393–402.

R. Roll, (1977a) "A Critique of the Asset Pricing Theory's Tests," *J. Finan. Econ.,* Mar. 1977, *4,* 129–76.

______, (1977b) "Testing a Portfolio for Ex Ante Mean/Variance Efficiency," work. paper, Grad. Sch. Bus., Univ. California-Los Angeles 1977.

W. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *J. Finance,* Sept. 1964, *19,* 425–42.

Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Reply

By ROBERT W. FOGEL AND STANLEY L. ENGERMAN*

The comments by Paul David and others in the March 1979 issue of this *Review* dealt mainly with technical issues related to the measurement of the relative efficiency of slave agriculture, and so does this reply.[1] The debate over these technical issues should not, however, obscure the marked shift in thought about the nature of the slave economy that has occurred during the past two decades. Until the mid-1950's it was widely believed that the slave plantations were unprofitable and inefficient enterprises that were kept in operation by a class prepared to sacrifice its private economic interest, enduring economic stagnation for the South, in order to maintain its political and cultural hegemony.

The first critical blow to this thesis was delivered by the now famous 1958 essay of Alfred Conrad and John Meyer. Marshalling the limited evidence available at that time, and using a standard capital model to estimate the internal rate of return on an investment in slaves, they concluded that rates of return on this form of capital compared favorably with rates on alternative investments. This finding set off an extended debate that took more than a decade to resolve. Eventually, refinements in the original capital model and a considerable expansion of the data base, including the Parker-Gallman sample of over 5,000 southern farms and large samples of slave prices and hire rates from probate and other records, placed the rate of return in the 6–10 percent range, thus demonstrating the allocative efficiency of the slave economy. The second blow to the traditional thesis was delivered by Richard Easterlin's development of regional income estimates extending back to 1840. Both these initial estimates and the subsequent refinements in them (see our 1971a paper) showed that the South experienced a high rate of growth in per capita income between 1840 and 1860, thus contradicting the view that the antebellum South was economically stagnant.

These twin blows to the traditional interpretation shifted cliometric attention to the question of the relative technical efficiency of slave and free agriculture in input utilization. While it is obviously premature to declare this question settled, a consensus has emerged on one critical point. It is now clear that the previous view that slave agriculture was less efficient than free agriculture is incorrect. What remains to be resolved is the exact margin of the advantage enjoyed by slave plantations, and the explanation for this margin.

I. Reply to David and Temin

A. *Measures of Land Inputs*

In their 1979 comment, Paul David and Peter Temin present a novel argument on measurement of the land input, which is quite different from their original one. They originally argued that because the northern transportation system was superior to the southern, the locational rent was a larger component of the value of northern farm land than of southern farm land (1974, pp. 772–73). They agreed that one could not merely measure the land input by acres in farms, that adjustments were needed for the fact that more capital was invested in land improvement in the North than in the South, as well as to take account of differences in natural fertility. But they objected to our use of land values for this purpose. Since we had measured output in uniform national prices, they argued that the procedure employed in our 1971 and 1974 computations failed to eliminate the locational component of land values, and hence biased the result in favor of the South.

---

*Harvard University and University of Rochester, respectively. Research on this paper was supported by National Science Foundation grants SOC 76-002 and GS-27262, and a grant from Harvard University. We have benefited from comments and criticisms by Fred Dong, Richard Freeman, David Galenson, Zvi Griliches, Thomas F. Heurtas, Mark Hopkins, Laurence Kotlikoff, Frank Lewis, Donald McCloskey, Paul F. McGouldrick, Jacob Metzer, John Olson, Hugh Rockoff, Donald F. Schaefer, Mark D. Schmitz, Ralph Shlomowitz, Kenneth Sokoloff, and Gavin Wright.

[1] Because of space limitations, some of the issues raised by these comments, including those of Thomas Haskell, are dealt with elsewhere (see our forthcoming book).

The critical issue in this argument is the direction of the bias. David and Temin assumed that because the North had a better transportation system than the South, the locational component in the value of farm land was greater in the North than in the South. While such an inference would be warranted in a partial equilibrium model, it is not appropriate to the actual situation in the antebellum United States. Between 1840 and 1860 roughly 70 percent of some 27,000 miles of new railroad track was laid in the North (see George Taylor, p. 79), bringing millions of acres close to a rail route. The massive increase in supramarginal land brought about by the railroad construction probably lowered rather than raised the average locational rent per ace of land in farms. The regressions reported in our 1977 paper confirmed that the locational component, on both a per acre and a per farm basis, represented a larger share of land values in the South than in the North. Consequently the effect of removing the locational component of land values was to raise the southern index of total factor productivity (G_s) relative to the northern index (G_n), rather than reduce the ratio as David and Temin had asserted (symbols are defined in Table 1).

TABLE 1—DEFINITIONS OF PRINCIPAL SYMBOLS

G_i = geometric index of total factor productivity for farms of a given category =

$$\frac{Q_i}{L_i^{\alpha_L} K_i^{\alpha_K} T_i^{\alpha_T}}$$

G_s/G_n = the geometric index of southern relative to northern total factor productivity =

$$\frac{Q_s/Q_n}{(L_s/L_n)^{\alpha_L}(K_s/K_n)^{\alpha_K}(T_s/T_n)^{\alpha_T}}$$

Q = output
L = input of labor
K = input of capital
T = input of land
α_L, α_K, α_T = shares in the value of output of labor, capital, and land
V = value of land plus improvements
I = number of improved acres
U = number of unimproved acres
γ_1 = price per acre of improved land (including improvements)
γ_2 = price per acre of unimproved land
N = the output point of a normalized northern farm
P = the price of output (cotton or other commodity, depending on context)
S_i = the output point of a normalized southern farm of the ith size class
R_i = the revenue of a normalized farm of the ith category
A_c = the efficiency index for cotton output
A_o = the efficiency index for all noncotton output
Y = the value of crop output
Y_c = value of cotton output
Y_o = value of all crop output except cotton
X = dummy variable taking the value 1 for farms with 16 or more slaves
A = the intercept of a Cobb-Douglas production function
Z = a geometric average of inputs
α_i = output elasticities of inputs
σ = the scale coefficient ($\sigma = \Sigma\alpha_i - 1$)
s = subscript denoting South
n = a subscript denoting North
g = a subscript denoting slave farms
f = a subscript denoting free farms

In their 1979 comment David and Temin not only tacitly accepted our finding that capitalized locational rents per acre were larger in the South than in the North, but they simultaneously redefined the land input, making it equal to capitalized locational rents, and nothing more. Their algebraic proof that we eliminated "land" from the production function turns purely on their tacit redefinition of the land input. They now argue that instead of removing capitalized locational rents, we should raise the per acre value of these rents on the less well-situated farms to the average level ($12.61) of those on the best-situated farms.

The locational component of the value of an improved acre is, of course, not a mea-

TABLE 2—THE EFFECT OF VARIOUS ADJUSTMENTS FOR THE QUALITY OF LAND ON THE SOUTHERN INDEXES OF TOTAL FACTOR PRODUCTIVITY
($G_0 = 100$)

Size of Farm as Measured by the Number of Slaves	T_i/T_0 Measured by Value of Land (V) (F-E, 1974) (1)	T_i/T_0 Measured by the Value of Improvements Plus Differential Natural Fertility $(\gamma_2-\gamma_1)I$ (F-E, 1977) (2)	T_i/T_0 Measured by Land Values, with an Upward Valuation of the Locational Component on Improved Land Only $V+I(12.61-\gamma_2)$ (D-T modified by F-E) (3)	T_i/T_0 Measured by Land Values, with an Upward Valuation of the Locational Component on Both Improved and Unimproved Land $V+(I+U)(12.61-\gamma_2)$ (D-T, 1979) (4)
0	100.0	100.0	100.0	100.0
1–15	107.7	100.8	109.7	113.7
16–50	144.7	133.1	148.5	159.8
51+	133.5	147.7	147.3	174.6
All Slave Farms	128.5	125.0	134.1	146.8
All Farms	123.2	120.4	127.9	138.2
Farms Not Using the Gang System	104.8	100.5	106.0	108.5
Farms Using the Gang System	140.3	138.8	148.0	165.6

Sources: Computed from data in our 1977 paper, Table 6, and in the Parker-Gallman sample.

Note: \$12.61 is the value of γ_2 for plantations in 51+ category, and G_0 is the index of total factor productivity on southern farms with zero slaves.

sure of an input into the production of farmgate crops. Quite the contrary, it measures the saving in a particular input, which is not land but transportation. Moreover, this transportation input is not used to produce farmgate crops, but to deliver them to market once they are produced. Since the locational component of land is not an argument of the farmgate production function, the land input is properly measured by improvements plus superior natural fertility. Insofar as the production of crops is concerned, it makes little sense to treat improvements to the land as if they were less integral to the land input than natural fertility. Most of the value of the improved acres in farms in 1860 was due not to pure fertility differentials but to improvements (see our 1977 paper, fn. 19).

While a case could be made for inflating the value of the locational component of improved land, there is no economic justification for extending that procedure to unimproved land, as David and Temin do. In 1860, 46 percent of the acreage of northern farms and 70 percent of the acreage of southern farms was unimproved (see U.S. Bureau of the Census, 1895, pp. 84–100). Such land was not an input into the current production of crops, but was kept in inventory by farmers who expected to improve it at some future date or sell it to someone else who would. It had value not because of its contribution to current output, but because of a potential contribution that would be realized at some future date.[2]

Column (4) of Table 2 displays the set of southern indexes obtained with the upward valuation of locational rents now advocated by David and Temin (D-T, 1979); column (3) displays the indexes when the inflation of locational rents is applied only to improved land (i.e., excluding land held in inventory for future use in crop production); and column (1) displays the set originally published in *Time on the Cross* (F-E, 1974).

---

[2] It could be argued that a small fraction of γ_2 should be retained as a measure of the input of grass and mast into the raising of livestock. Such an adjustment would have little impact on either the North-South or the intra-South comparisons of total factor productivity. It cannot, in any event, be used to justify the multiplication of γ_2 by four or five times, as David and Temin have done.

TABLE 3—THE EFFECT OF VARIOUS ADJUSTMENTS FOR THE QUALITY OF LAND ON THE NORTH-SOUTH COMPARISON OF TOTAL FACTOR PRODUCTIVITY ($G_n = 100$)

	Values of the Index of Total Factor Productivity for Various Classes of Farms in the Parker-Gallman Sample			
Alternative Measures of the Ratio of Land Inputs	Farms Not Using the Gang System (1)	All Farms (2)	All Slave Farms (3)	Slave Farms Using the Gang System (4)
1. T_i/T_n Measured by the Value of Land (V)	114.5	134.7	140.5	153.4
2. T_i/T_n Measured by the Value of Improvements plus Differential Fertility $(\gamma_2-\gamma_1)I$	123.6	148.1	153.8	170.7
3. T_i/T_n Measured by Land Values, with an Upward Valuation of the Locational Component on Improved Land only $V+I(12.61-\gamma_2)$	112.1	135.3	141.8	156.5
4. T_i/T_n measured by Land Values, with an Upward Valuation of the Locational Component on both Improved and Unimproved Land $V+(I+U)(12.61-\gamma_2)$	101.7	129.5	137.6	155.2

Sources: From our 1974 book, Vol. II, Table B.23; our 1977 paper, Table 6; Table 2 above; and the Parker-Gallman sample.

As can be seen from a comparison of columns (1) and (4), David and Temin's procedure increases the productivity advantage of the intermediate and large slave farms relative to either free southern farms or small slave farms (1–15 slaves). Under the David and Temin procedure, the slave farms are 46.8 percent more efficient than free southern farms; and the slave farms using the gang system (those with 16 or more slaves) are 52.6 percent more efficient than the farms which did not (those with 0–15 slaves). Consequently, far from overturning our results, the procedure advocated by David and Temin in their comment actually reinforces them. But even if one accepts their rationale for inflating locational rents, the correct implementation of this procedure is, as we have argued, the set of indexes displayed in column (3). This computation gives virtually the same edge to the gang system ($148.0 \div 1.060 = 139.6$) as obtained from the set of indexes we prefer ($138.8 \div 1.005 = 138.1$) and which are displayed in column (2).

What of the North-South comparison? Although David and Temin increased the locational component of land values on free southern farms to $12.61 per acre, they failed to make a similar adjustment for northern farms. When symmetry is maintained, the procedure they now advocate does little to diminish the advantage of slave farms over the free farms of the North. As column (3) of Table 3 shows, this procedure still gives slave farms an edge of 37.6 percent (line 4). Moreover, the edge of farms using the gang system (see the entries in column (4)) is actually higher under their procedure than under the one reported in *Time on the Cross*. A comparison of lines 1 and 4 of Table 3 reveals that only the efficiency of nongang farms is significantly diminished relative to that of northern farms. This reduction occurs because 75 percent of the acreage of free southern farms was unimproved, while less than half of the acreage of northern farms was unimproved (see our 1977 paper, Table 6 and fn. 19). It follows that the inclusion of unimproved land (with greatly inflated price weights) in the production functions introduces a substantially greater downward bias in the measured efficiency of free southern farms than of northern farms. When the inflation of the land rent is confined to improved land (see line 3) the results are quite similar to our original estimates (see line 1). Thus David and Temin obtained sharp drops in the efficiency of southern free farms relative to northern farms because their computations involved two inappropriate procedures: one was the failure to allow for a transportation subsidy to northern farmers corresponding to that presumed for southern farmers; the other was the inclusion of a greatly inflated locational rent on unimproved land that was not used, or used to a negligible degree, for purposes of current production.[3]

B. *The Product-Mix Problem*

In their 1974 paper David and Temin developed a model showing that despite differences in the indexes of total factor productivity, slave and free farms could be located on the same production-possibility frontier. That model, which was reiterated in their 1979 comment, is replicated in Figure 1a, with a slight modification in symbols to avoid confusion between their notation and ours.[4] The curved line running through points N and S is the production-possibility frontier. Point S represents the southern output of cotton and "grain" (which stands for the output of noncotton crops and livestock). The point N gives the output of cotton and grain for the North. Outputs of both sectors are normalized by expressing them as ratios with respect to the average unit of inputs. Placing both N and S on the same production-possibility curve implies that they have the same index of technical efficiency. The line RR, which is tangent to the production-possibility curve, represents the total revenue of the output of a normalized southern farm when valued at the prevailing prices of cotton (P_c) and grain (P_g) with the ratio P_c/P_g given by the slope of RR. The distance of RR from the origin along any ray may be taken to represent the southern index of total factor productivity. The dashed line which intersects the production-possibility curve at N is the total revenue of the output of a normalized northern farm when valued at the indicated prices. If the distances from the origin of the two revenue curves are measured along the grain axis, then

$$Q_{gs}/Q_{gn}=G_s/G_n$$

Thus David and Temin claimed to have shown that the deviation of our index of relative total factor productivity (G_s/G_n) from 100 is nothing but a price effect, and has no implication regarding technical efficiency.

Whatever merit there is to David and Temin's argument, it is untrue that extreme

---

[3] In our 1977 paper we shifted to the indexes which measure the land input by $(\gamma_1-\gamma_2)I$ for the purposes of the intra-South efficiency comparisons (i.e., the indexes of column (2) in Table 2), but we continued to use the indexes presented in Table B.23 of *Time on the Cross* for the North-South comparisons. We did so because other adjustments in the northern efficiency index, still not fully refined, tend to offset the effect of the North-South adjustment for locational rents. These include adjustments for omitted products (principally firewood, which was a more important part of northern than southern farm output) and for regional differentials in mortality rates. While the lower northern mortality rate has, of course, been noted, its implications for the measurement of productivity have not been pursued. Preliminary estimates indicate that the mortality and other adjustments could raise the values of G_n, after correction for locational rents, by as much as 10 percent. One can gauge the effect on the North-South comparison by dividing the entries in line 2 of Table 3 by 1.1. The resulting figures (112.4, 134.6, 139.8, and 155.2) are quite close to the corresponding entries of line 1 of the same table. The full set of adjustments to G_n will be reported in a forthcoming study.

[4] It should be noted that in their 1979 comment David and Temin interchanged the labels of the x and y axes. Figure 1a conforms to the diagram in their 1974 essay.

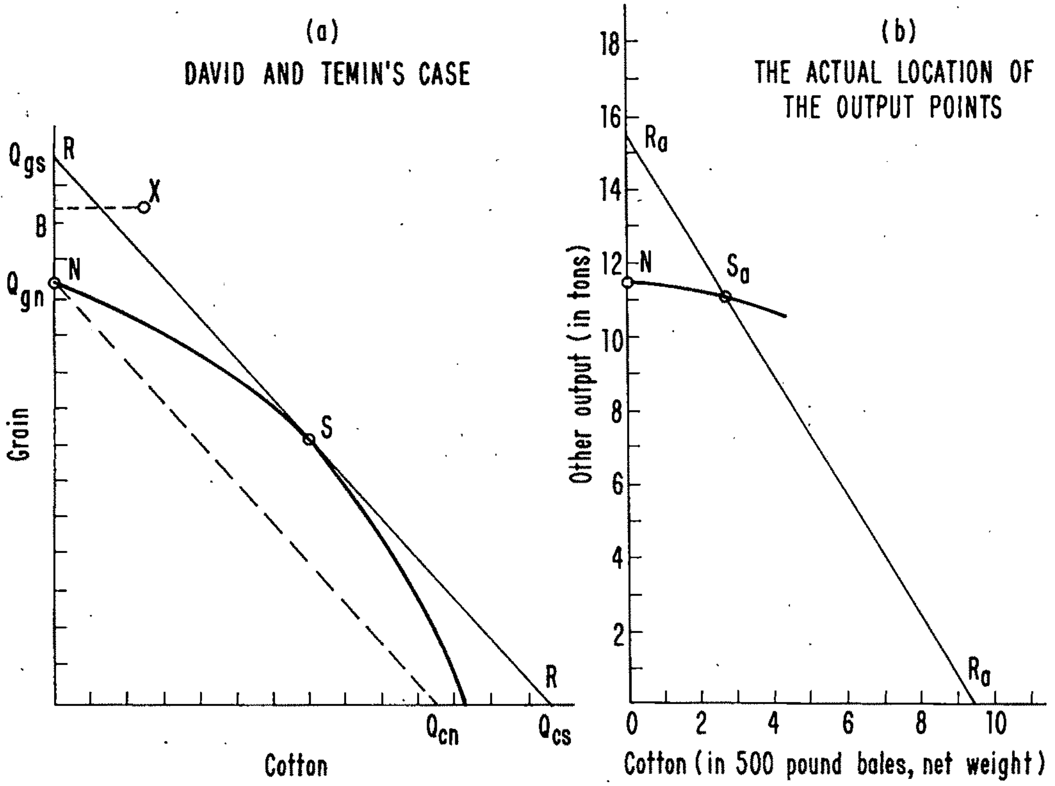


FIGURE 1

differences in product mix necessarily prevent inferences regarding the relative technical efficiency of northern and southern agriculture. If, in Figure 1a, *S* were located anywhere to the northeast of *N*, then southern agriculture would be unambiguously more efficient than northern agriculture. If, for example, southern output was at point *X*, then the efficiency of southern agriculture would exceed that of northern agriculture by at least the ratio BN/NO. That ratio is not, however, a greatest lower bound for it is based on the false assumption that the marginal rate of substitution at *X* between grain and cotton was zero. A greater lower bound on the southern efficiency advantage could be obtained by allowing for the largest plausible deviation of the price ratio from the marginal rate of substitution at point *X*.

The purpose of the preceding example was to demonstrate that David and Temin's "inescapable" conclusion is not entailed by theory. It depends on the positions at which they located *N* and *S*, and the slope that they assigned to the revenue line. What is at issue, then, are not points of logic but the validity of the empirical assumptions implicitly introduced into their diagram. They assumed that the points were so arranged that they might well lie on the same concave production-possibilities curve. We do not dispute their location of the northern output (*N*) on the grain axis—although in 1860 two northern states, Missouri and Illinois, produced small quantities of cotton (see U.S. Bureau of the Census, 1862, pp. 200–01). But they erred in locating the southern output so far to the south and east of *N*. Figure 1b shows that southern output in 1860 was actually located at S_a. While this point is not in the unambiguous zone, it is nearly so, lying almost due east of the northern output point. It follows that *N* and S_a could lie on the same concave production-pos-

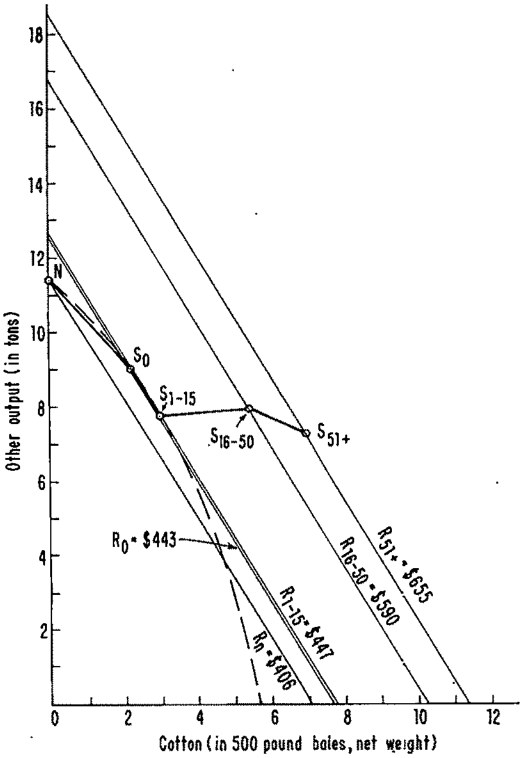


FIGURE 2. ACTUAL OUTPUT POINTS FOR FARMS OF VARIOUS SIZE CLASSES IN THE COTTON BELT

sibility frontier, *and be tangent to the price line*, only if that frontier is kinked, or nearly kinked. Figure 1b shows that these points cannot lie on a more normally shaped frontier, such as that shown by the heavy line, since its slope at point S_a is quite flat and so implies a ratio of cotton to other farm produce prices well below that actually observed during any year of the nineteenth century. A kinked, or nearly kinked, frontier is ruled out by the wide variations in the noncotton to cotton output ratio across farm-size classes during given years, and substantial shifts in this ratio from year to year, associated with relatively small variations in the ratio of product prices.[5] David and Temin also misjudged the relative prices of cotton and grain. The appropriate slope of the revenue curve, given by the line R_aR_a, is 50 percent greater than they assumed.

The notion that the northern farms, the free southern farms, and the various classes of slave farms all lie on the sample production-possibilities curve contains a modicum of plausibility only when one aggregates over the output of all southern farms, ignoring the possibility that technical efficiency may differ by the size of the plantation. In Figure 2, the outputs for each class of farms are shown.[6] Also indicated are the normalized revenue curves associated with each point. The subscripts on R_i and S_i indicate the size class to which each revenue line and output point pertains. Thus $R_{16\text{-}50}$ is the normalized revenue line associated with farms having 16 to 50 slaves. The solid line passing through the five output points approximates the shape of the production-possibilities curve if all classes of farms were on it. Clearly it is impossible for all these points to be on a smooth, concave (from below) production possibility curve. It is plausible, however, that N, S_0, and $S_{1\text{-}15}$ lie on such a curve as is shown by the dashed line running through these points. But if the free farms of the North and South, as well as the small slave farms of the South, all had the same basic production technology then one could adjust for the North's inability to produce cotton by comparing the normalized revenue line through either S_0 or $S_{1\text{-}15}$ with the normalized revenue lines through $S_{16\text{-}50}$ and S_{51}.

In their comment David and Temin argue that differences in the relative prices facing the different classes of southern farms could explain their different locations on the same production function. The implausibility of that conjecture for all classes of farms except those in the zero and 1–15 slave classes

---

[5] This point is elaborated below in the discussion of Figure 3, and in our forthcoming book.

[6] It should be remembered that indexes of total factor productivity by farm size within the South were computed from the Parker-Gallman sample, which is limited to the counties of the South that specialized in the production of cotton. Since southern counties specializing in grain and meat products were omitted, the amount of "other output" in each of the southern size classes shown in Figure 2 falls below that shown for the South as a whole in Figure 1b.

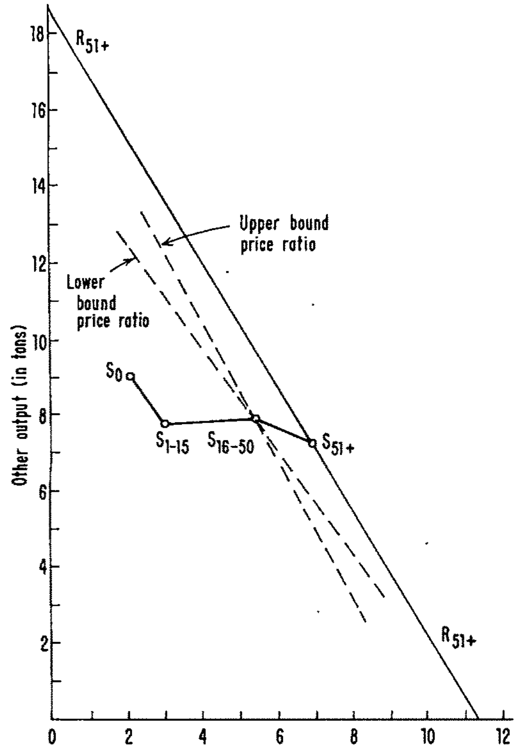


FIGURE 3. UPPER AND LOWER BOUNDS ON THE PRICE RATIO FACING INTERMEDIATE-SIZE PLANTATIONS RELATIVE TO THE PRICE RATIO FACING LARGE PLANTATIONS

is demonstrated in Figure 3. The total revenue line marked R_{51+} shows the ratio of P_c/P_o at national average prices in 1860.[7] The slopes of the straight lines connecting the various output points are upper bounds, at the higher of each pair of points, on the absolute value of the slope of a concave curve passing through both points.[8] Thus comparison of the slope of the line segment $S_{51+}S_{16\text{-}50}$ with the slope of R_{51+} implies that the ratio of P_c and P_o facing intermediate plantations was only 0.26 as high as that facing plantations with 51 or more slaves. Since large and intermediate plantations coexisted in the same counties and shipped to the same markets, such a large differential is, on the face of it, implausible. Indeed, it is possible to establish upper and lower bounds on the slope of the price line at $S_{16\text{-}50}$, expressed as a proportion of the slope of R_{51+}. These bounds are based on the estimate of the average value of the locational advantage per acre of large plantations and on the average output per acre of corn and cotton. The results of the computation are also displayed in Figure 3, which shows that a price line with a slope lying between 0.83 and 1.10 of that facing the large plantations bounds the range of plausible alternatives.[9]

---

[7] The slope of this line is 1.63. If we had used the ratio of the average price of cotton to the average price of corn at New Orleans for 1860, the slope, computed from data in Arthur Cole, would have been 1.89.

[8] For example, the absolute value of the slope of the line connecting S_{51+} and $S_{16\text{-}50}$ is 0.43. The parabola passing through both points and constrained to have a slope of -0.43 at S_{51+} is given by $Q_o = 10.233 - 0.4284Q_c - 0.00010Q_c^2$. The absolute value of the slope of this curve at $S_{16\text{-}50}$ is 0.4296.

[9] The average value of the locational advantage of large plantations over smaller slave and free farms was about \$10 per acre (see our 1977 paper, Table 6). Given a discount of 0.1 (see our 1974 book, Vol. II, p. 78), it follows that the extra cost of shipping the output of an acre to market for the less advantageously located intermediate plantations was about \$1.00. Since the value of the average output of an acre of cotton was \$20.36 for large plantations (using a yield of 177.1 pounds per acre and a price of \$57.5 per net bale of 500 pounds), the value of the same output on smaller plantations was \$19.36. Thus the farmgate price of cotton to intermediate plantations was 19.36/20.36 or 0.95 of that facing plantations in the 51+ category. Using a similar procedure, with a yield of 840 pounds per acre and a price of \$16.43 per ton, the farmgate price of corn to intermediate plantations that produced a surplus of this crop was 5.90/6.90 = 0.86 of the price facing large plantations. It follows that the price ratio to intermediate plantations shipping both crops, relative to the price ratio facing large plantations shipping both crops, was 0.95/0.86 = 1.10, or 10 percent higher. Of course some intermediate plantations that shipped cotton ran deficits of corn. Such plantations would evaluate the price of corn at what it cost delivered to the farm. Consequently, the farmgate price of corn to them was not 5.90/6.90 but 7.90/6.90 = 1.14 of the price facing large plantations. It follows that the price ratio to an intermediate plantation shipping cotton but buying corn, relative to the price ratio facing the large plantation, was 0.95/1.14 = 0.83. It should be noted that these calculations are based on the assumption that both sizes of farms had land of average fertility. A less well-located farm with land especially suited for the production of corn, or other crops with a relatively low value per pound, might export relatively more corn

Although David and Temin's objections have been cast largely in theoretical and procedural terms, they are at the root disagreements over points of fact. David and Temin assumed that the work year was at least 10 percent longer in the South than in the North, although the evidence indicates the reverse. They assumed that locational rents were a larger component of northern than southern land values, although locational rents per acre were more than twice as high in the South as in the North. They assumed that the South's production of noncotton agricultual output (grain) per normalized farm was barely half the level that prevailed in the North, although as Figure 1 shows, it is 96 percent. This last point is perhaps the most critical of their factual errors. A normalized southern farm produced nearly as much grain as its northern counterpart with only about 80 percent of its supply of improved land and about 85 percent of its supply of labor.[10] We will return to this point again in the discussion of Wright's comment.

II. Reply to Wright

Gavin Wright's position (1975, 1976, 1979) is somewhat different from that of David and Temin. He acknowledges the existence of productivity differentials between slave and free farms, but denies that these were related to the gang system of slave labor employed on the intermediate (16–50 slaves) and large plantations. He postulates that slave and free farms were equally efficient in cotton production; that they were also equally efficient in noncotton production; but that they were more efficient in producing cotton than noncotton products. He proceeds directly from this postulate to two conclusions: 1) that the higher level of G_i/G_0 (where G_0 is the efficiency index on free southern farms) is due to the fact that cotton represented a higher share of output on slave than free farms; and 2) that if the free farms had been willing to increase the share of cotton in their total output to the level of large slave plantations, they would have had the same efficiency index.

In his 1979 comment Wright dismissed our demonstration that his postulate implies that free farmers were willing to pay an extraordinarily high "insurance premium" (close to half their annual income) in order to obtain, at best, a small reduction in the variance of income. He charges that we "forget completely about the exceptional character of the year in question" (p. 224), (1859–60), but he does not provide an estimate of what the insurance premium would have been under what he contends were normal prices and yields. If we reduce cotton income, using Wright's prices (see his fn. 1, 1979) and the midpoint of his interval on excessive yields, the annual insurance premium falls from the 48 percent of mean income that we estimated to 33 percent. If we also accept his contention that corn yields were down, say by 10 percent, and so raise corn income appropriately, the insurance premium falls to 29 percent. Thus, even if Wright's contentions about the abnormalities of 1860 yields and prices are granted, his postulate implies that free farmers were willing to forego more than one-quarter of their annual income for insurance that barely reduced, and may even have increased, the coefficient of variation (see our 1977 paper, p. 290, especially fns. 28 and 30).[11]

---

than a better located farm. We focused on the comparison between large and intermediate plantations because, with the previously noted exception of S_0 and $S_{1\text{-}15}$, that is the case most favorable to David and Temin's contention that local differences in relative prices explained the difference in product mix of farms sharing a common production frontier. The slope of the line connecting S_{51+} and S_0 is even less than that connecting S_{51+} and $S_{16\text{-}50}$. And since the slope of the line connecting $S_{16\text{-}50}$ and $S_{1\text{-}15}$ is positive, the farms represented by these points could share a common frontier only if one of the farmgate prices facing farms with 1–15 slaves was negative.

[10] The number of improved acres devoted to cotton in 1859–60 was estimated by dividing 177.1 pounds into the census figure for cotton output, with bales taken at 400 pounds net weight. The share of the labor force devoted to cotton is based on the data in our 1977 paper and forthcoming book.

[11] Because of limits of space we are unable to deal with Wright's conjecture that unusually favorable weather made 1859 per acre cotton yields so high that

The burden of Wright's argument on the crop-mix issue turns on a single equation relating output per hand to a measure of scale and the cotton share of output. This equation, estimated in three versions for his comment (his equations (1)–(3)), was introduced in his 1975 critique and reemphasized in the 1976 critique. He argues that these regressions show not merely that total factor productivity and the cotton share were "closely linked," but that "crop mix called the tune" (1979, pp. 222–23).

Even if there were no theoretical points at issue, this claim would be questionable because Wright's productivity measure is not an index of total factor productivity but a biased proxy for it. Some bias is introduced into his productivity ratio because it excludes noncrop output from the numerator; some because it excludes capital and land inputs from the denominator. But these biases are small compared to that introduced by the way in which the labor input was measured.

Since slave women and children were, on average, less productive than prime-aged slave men, it is necessary to convert persons of different ages and sexes into equivalent "full hands."[12] While Wright has recognized that in principle the labor should be measured in equivalent full hands, his measure does not have an appropriate adjustment for the labor of women and children. In the case of free farms, he agreed with Robert Russel that "white women and girls of small-farm families also worked in the fields to a considerable extent" (1969, p. 127; see Russel, p. 115), but nevertheless contended that the assignment of a zero weight to all females and to males 10–14 would not seriously bias his productivity measure. In the case of slaves, he assumed that women were as productive as males and thus assigned all women age 15–60 the same weight as prime-aged males. Quoting F. Linden (p. 166), Wright justified this procedure on the ground that "large and small plantations tended to be homogenous as to the sex and age ratios of their slave forces" (1969, p. 126). This statement is not borne out by the Parker-Gallman sample, which shows that

---

they "markedly effect" the efficiency calculation (1979, p. 220). The evidence against that conjecture is presented in our forthcoming book. It is also shown that even if one accepts Wright's most recent estimate of the magnitude of the random component of yields, the required adjustment still leaves the South with an efficiency advantage of at least 25.1 percent. More fundamentally, we object to Wright's statement that "numerous testimonials" to extraordinarily high per acre yields "are in fact available" and to his use of quotations from Ezekiel Donnell and the *American Cotton Planter* to support that claim. We do not agree that his position is supported by Donnell, who "cites the 'fine and very favorable picking weather' as the explanation for the cotton receipts 'larger than ever before known' in December 1859" (1979, p. 222). Nowhere in the five pages that Donnell devotes to the 1859 crop does he distinguish between per acre yields and an increase in acreage as the cause for the size of that crop. The phrases quoted by Wright come not from a discussion of output and yields, but from notes to a table of cotton prices and commercial transactions. The following is the full note with the abstracted phrases shown in italics:

> No very remarkable event occurred this crop year, and prices varied but little, its entire course. Early in September, the free receipts at the ports caused a dull, heavy feeling here and abroad, which continued through October. About the 1st of November, accounts of frost stiffened the market temporarily, but with subsequent *fine and very favorable picking weather*, prices soon fell off again. Owing to a rise in the Southern rivers, the receipts for a time in December were *larger than ever before known*, but under an improved demand, prices were sustained. [p. 496]

As we read this note, the unusual December receipts are attributed not to unusually high acre per yields, but to unusually good conditions of river transportation that temporarily permitted a larger portion of the crop to be delivered during December than was normal. To interpret it as evidence of extraordinarily high per acre yields is to add something that is not there. Wright's quotations from the *American Cotton Planter* also fail to address the question of per acre yields vs. a shift in resources and again focuses primarily on the state of the market. However, the September 1859 issue of the same journal contains an unambiguous report on the shift of land from corn to cotton between the 1858 and 1859 crop years: "The fair price of cotton last winter, induced planters this year to plant a greater breadth of land in cotton than usual, to the neglect of corn; trusting to the West to supply them with both corn and pork" (p. 272).

[12] See our 1977 paper, Table 1, for a brief description of our procedure in making these age and sex adjustments.

TABLE 4—THE RATIO OF WRIGHT'S MEASURE OF LABOR INPUT TO A MEASURE OF EQUIVALENT FULL HANDS

Size Class as Measured by the Number of Slaves Per Farm (1)	Average Number of Equivalent Full Hands Per Farm (2)	Average Number of Hands Per Farm under Wright's Procedure (3)	Ratio of Column (3) to Column (2) (4)
0	2.31	1.65	0.71
1–15	4.66	4.60	0.99
16–50	12.43	15.19	1.22
51+	37.52	48.27	1.29

Source: Computed from the Parker-Gallman samples in the manner indicated by Wright 1976, p. 334; and our 1977 paper, Table 1.

Note: The number of farms was 4,820. Of the 5,228 farms in the Parker-Gallman sample, 408 were omitted because of zero entries in one or more of the relevant variables.

the age and sex structure of the slave population varied with plantation size. The net effect of Wright's assumptions was to underestimate substantially the labor input on free farms and to overestimate the labor input on slave farms, the extent of the overestimating rising with the size of the slave farm. Table 4 shows that with the exception of small slave farms, where the underestimate of the labor input of free women and children virtually offsets the overestimate of labor input of slave women, this error is in excess of 20 percent. The systematic mismeasurement of the labor input, by itself, is sufficient to explain why Wright's regressions yielded negative or statistically insignificant scale coefficients.[13]

But suppose Wright had measured the labor input correctly. Would the positive correlation between his productivity index (Y/L) and the cotton share (Y_c/Y) constitute evidence in favor of his hypothesis and against ours?[14] If it is true that the gang system gave larger plantations a comparative advantage in cotton production, then it follows that the cotton share should have been greater on large slave farms than on small free farms. Clearly, both hypotheses imply that the cotton share and the productivity index should be positively correlated. What distinguishes them is neither the sign nor the strength of the correlation, but the direction of causation. We argue that the cotton share increased because some farms were relatively more efficient in cotton production than others. Since Wright holds that the ratio of the efficiency indexes in cotton and noncotton production (A_c/A_o) was more or less equal on all types of farms and that $A_c > A_o$, he contends that it was variations in the cotton share that "called the tune." The two hypotheses are also distinguished in another respect. Wright's hypothesis implies that the productivity index was a continuous function of farm size because the cotton share rose across all size classes. We, on the other hand, argue that the relationship between G_i and farm size was discontinuous because there was a threshold level for the employment of the gang system. Nothing in Wright's regressions bear on the direction of causation; nor do the regressions test the hypothesis that the relationship between G_i and size was discontinuous. Fortunately, evidence bearing on both issues is available.

---

[13] Indeed, the bias in Wright's measure of the labor input is so large that it wipes out the entire differential in total factor productivity between free and large slave farms, making large slave farms 5 percent less efficient than free farms. See fn. 18 below.

[14] In this discussion we assume that Y/L and G are well correlated, as is the case. In other words, we accept an index of labor productivity (when the labor input is properly measured) as an adequate, though not a perfect, proxy for the total factor-productivity index.

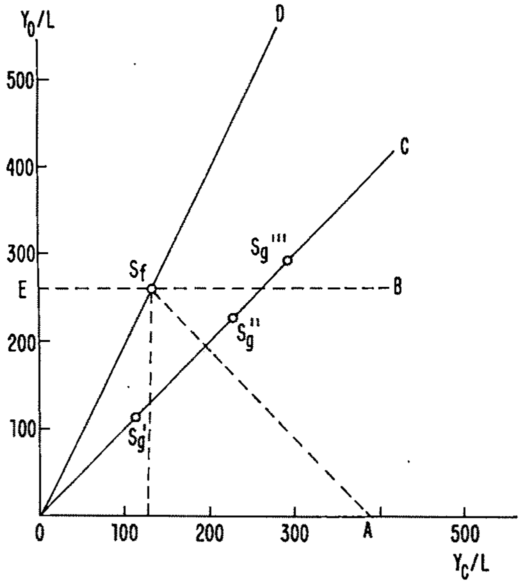


FIGURE 4. LOGICAL LIMITS OF WRIGHT'S INTERPRETATION OF THE CORRELATION BETWEEN OUTPUT PER HAND AND THE COTTON SHARE

Assessment of this evidence will be facilitated if we restate Wright's argument within the framework of a production space diagram, such as Figure 4. This figure is similar to the diagrams employed in our discussion of David and Temin's argument, but is modified to conform with Wright's variables after correcting the measurement of the labor input for the biases previously specified. Output is, therefore, measured by value rather than in physical terms, noncotton output (Y_o) is limited to noncotton crops, and farms are normalized on the labor input alone rather than on all inputs. The slope of a line from the origin to any S_i in an index of the cotton share, since

$$\frac{Y_c}{Y} = 1/((Y_o/Y_c)+1)$$

where Y_c/Y is the cotton share and Y_c is the dollar value of cotton output. Thus the smaller the slope of such a line, the larger the cotton share. Output points are again designated by the letter S. In order not to complicate this figure unduly, all slave farms are represented by a single category, with the subscript f used to designate free southern farms and the subscript g used to designate slave farms. In the subsequent figure, allowance will be made for productivity differentials between the various size classes of slave farms.

In Figure 4, the line $0D$ indicates the cotton share of free farms. S_g must be located along the line $0C$ which indicates the cotton share of slave plantations. The point S_g cannot be located to the left of the line S_fA,[15] as is S_g', since Y/L was greater for slave than free farms and all points to the left of that line have lower values of Y/L than free farms. If S_g was located at any of the points on $0C$ above the line EB, as is S_g''', then slave plantations would be unambiguously more efficient than free farms. Thus for Wright's interpretation of his regressions to be true, S_g must fall in the top half of the ambiguous zone—that is, on a point along $0C$ located within the area bounded by AS_fB, as in the case for S_g''.

Three conclusions follow from Figure 4. First, given that slave farms had both a higher income per worker (Y/L) and a higher cotton share (Y_c/Y), the positive correlation between Y/L and Y_c/Y rules out the possibility that slave farms were unambiguously *less* productive than free farms. Second, the most that Wright can claim about the location of S_g on the basis of this positive correlation is that it *might be* in the ambiguous zone. Third, in order to use regression analysis to demonstrate that S_g *is* in the ambiguous zone, it must be shown not only that Y/L and Y_c/Y are positively correlated, but also that Y_c/L and Y_0/L are negatively correlated.[16] For if both Y_0/L and Y_c/L are greater for slave than for free farms, S_g must lie in the zone bounded by BS_fD, in which case slave farms would be unambiguously more efficient than free farms.

---

[15] Since the value of Y/L at any point is given by the sum of the coordinates of that point, the line S_fA represents the locus of all points with the same Y/L as free farms and with at least as large a cotton share.

[16] Note that such a negative correlation would show only that S_g was in the ambiguous zone. It would not rule out our hypothesis or confirm Wright's.

What is critical then, is the relationship between Y_0/L and Y_c/L. Wright presents no direct evidence on the sign of this relationship, and refers instead to his safety-first hypothesis which involves the proposition that as farmers increased Y_c/L, they reduced Y_0/L; that is, that farms with relatively high cotton shares produced less food per equivalent hand than those with a relatively low one. To reason in this way is quite circular, since *both* Wright's crop-mix thesis and his safety-first thesis require a demonstration that there was a negative relationship between Y_c/L and Y_0/L.

Rather than presenting direct evidence on the relationship between Y_c/L and Y_0/L, Wright offers two indirect pieces of evidence. One is that the farm population per improved acre was greater on free farms than on slave farms.[17] But as is immediately apparent, this fact does not rule out the possibility that Y_0/L rose with Y_c/L. The other piece of evidence is his second regression, which is run on data for free farms only. Wright suggests that having eliminated the influence of the gang system, it is clearly crop mix rather than "true" differences in productivity which account for the positive correlation between Y/L and Y_c/Y. But is that inference so obviously warranted? While the gang system was one factor which affected the comparative advantage of farmers with respect to cotton, it was not the only one. Growing cotton was a complex task, requiring specific skills. The distribution of these skills varied at least as much among free farmers as among slave-owning farmers. Since some free farmers were more adept than others in growing cotton, comparative advantage could explain the positive correlation between Y/L and Y_c/Y, not only in his first regression but also in his second one—although the explanation for the comparative advantage would be somewhat different in the two cases.

Figure 5 presents direct evidence on the relationship between Y_c/L and Y_0/L. Part (a) pertains to the case of free farms only. Here the farms are arrayed according to the ranking of their index of labor productivity. Thus S_1 represents the average normalized output point for the free farmers belonging to the lowest quartile of the index, while S_4 represents the corresponding output point for farmers in the highest quartile of the index. Since S_4 is to the northeast of S_3, and so on, there is no doubt that farms with the higher productivity indexes were more efficient than those with the lower ones. Wright's contention that the cotton share increased because farmers reduced noncotton output is clearly wrong. The farms with the higher productivity index grew more of both crops per hand. The cotton share rose with the productivity index because the increase in cotton output was greater than the increase in noncotton output. Figure 5b compares free and slave farms, with the subscripts on the S_i referring to the number of slaves. Since $S_{1\text{-}15}$ and $S_{16\text{-}50}$ are to the northeast of S_0, these points represent unambiguously higher levels of efficiency. Only S_{51+} is in the ambiguous zone. But as demonstrated in Figure 3 above, given the price ratios, it is quite likely that this point also represents a higher level of true productivity than S_0.

We are now in a position to show why Wright's equation (1) produces a statistically insignificant scale coefficient, even when the labor input is properly measured. The crux of the issue is that this equation compares the way in which the productivity index varies with size, when the cotton share is held constant. But as Figure 5 demonstrates, farms with a high cotton share not only had a higher productivity index but were actually more efficient than farms with a low cotton share, both across size classes and within size classes. Consequently, holding the cotton share constant, at say 50 percent, forces a comparison between free farms that are *more* efficient than the average of their

---

[17] Wright states that the ratio of persons per improved acre was nearly twice as great on free farms as on farms with 51 or more slaves. But this figure is biased upward by about 46 percent because Wright gave all farms within each size class equal weight. To obtain an unbiased estimate of the average number of persons per improved acre for each size class, persons per improved acre on each farm has to be weighted by the farm's share in the total improved acreage of its class.

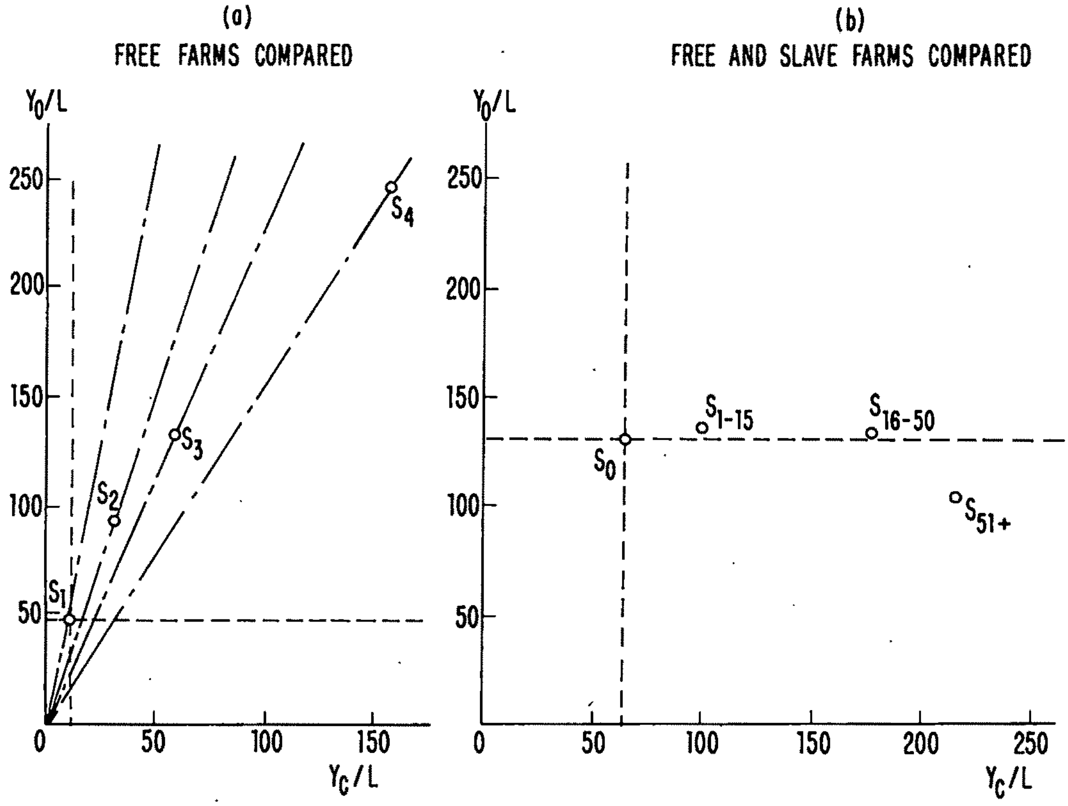


FIGURE 5. THE RELATIONSHIP BETWEEN Y_o/L AND Y_c/L

class and large slave farms that are *less* efficient than the average of their class. In other words, holding the cotton share constant is the econometric equivalent of excluding from the comparison both the *less* efficient of the free farms and the *more* efficient of the large farms. This point is demonstrated in Table 5, where the cotton share is held constant at 50±10 percent. That restriction leads to a comparison between the productivity index of free farms belonging to the seventh decile of the zero size class and the productivity index of large farms belonging to the fourth decile of the 51+ size class.[18]

A final point on Wright's regressions. We have argued that the gang system was one (but not the only) factor that made some farms more efficient in cotton production than others. Whatever the reason for having a comparative advantage in this crop, those farms that had it would have had a larger share of output in cotton than those that did not. Is it possible, within the framework of regression analysis, to separate out the effect of the gang system from other factors

---

[18] In his 1969 study Wright ran regressions, not on the entire set of valid observations in the Parker-Gallman sample, but on a subsample "of farms for which the cotton-corn ratio was restricted to a narrow range" (p. 129). Since the regressions run on this restricted subsample yielded generally nonsignificant scale coefficients, Wright rejected the hypothesis that there were economies of scale on slave plantations. This conclusion, as we pointed out (1971b, p. 364), was not warranted since Wright's restriction on the cotton-corn ratio had removed from the original Parker-Gallman sample the *less* efficient of the small farms and the *more* efficient of the large farms. Wright's regressions in his 1979 comment thus reproduced by econometric means the error introduced into his 1969 study through sample selection bias.

TABLE 5—THE CLASS MEANS OF Y/L IN BOTH A SUBSAMPLE HAVING A COTTON SHARE OF 50±10 PERCENT AND IN THE COMPLETE SAMPLE

Size class as Measured by Number of Slaves (1)	Mean Values of Y/L in the Complete Sample ($) (2)	Mean Values of Y/L in the Subsample with a Cotton Share Of 50±10 Percent ($) (3)	Percentage Deviation of Subsample Mean from Sample Mean (4)	Location of the Subsample Mean in the Complete Within-Class Distribution of Y/L (5)
0	196	223	+14	bottom of 7th decile
1–15	235	251	7	middle of 6th decile
16–50	307	283	−8	top of 5th decile
51+	318	250	−21	top of 4th decile

Source: Parker-Gallman sample.

that affected the efficiency of production? Wright's attempt to do so failed because he compared the productivity of subsets of farms drawn from different deciles of their respective size-class distribution. Clearly, the comparison should turn on the means of the variables of given size classes. That is precisely the nature of the comparison when an ordinary Cobb-Douglas production function is estimated. Such a regression, run on the Parker-Gallman sample, yields

$$(1)\quad \hat{Q}-\hat{L}=\underset{}{2.8984}+\underset{(0.0113)}{0.1815}(\hat{K}-\hat{L})+\underset{(0.0125)}{0.2606}(\hat{T}-\hat{L})+\underset{(0.0124)}{0.0645}\hat{L}$$

$$R^2=0.70;\ N=4299$$

where a "hat" over a variable indicates the natural logarithm of that variable, and standard errors are shown in parentheses.

While the positive and statistically significant scale coefficient indicates that productivity increased with farm size, this regression also fails to separate out the influence of the gang system from other factors that might have varied continuously with size. Such separation is possible, however, because the gang system could not be effectively employed unless there were a minimum number of hands. Farms with many more hands than this minimum number—the threshold appears to have been about 10 hands or roughly 16 slaves—had multiple gangs. To separate the influence of the gang system per se from factors varying continuously with size, we ran equation (1) with a dummy variable (X) that took the value 1 for farms with 16 or more slaves:

$$(2)\quad \hat{Q}-\hat{L}=3.088+\underset{(0.011)}{0.177}(\hat{K}-\hat{L})+\underset{(0.112)}{0.247}(\hat{T}-\hat{L})-\underset{(0.018)}{0.059}\hat{L}+\underset{(0.039)}{0.397}X$$

$$R^2=0.71;\ N=4299$$

Not only is the coefficient on X large and statistically significant, but the scale coefficient becomes negative. This finding suggests that it was the gang system, and not the factors varying continuously with size, that gave the intermediate and large plantations their edge in productivity. It is consistent with the results obtained from the computation of total factor productivity indexes and the evidence derived from the study of the managerial documents of slave plantations (see our 1974 book, Vol. I, pp. 202–09; our 1977 paper, pp. 290–92; Jacob Metzer; John Olson). This evidence points to the gang system as the instrument which made it possible for slaveowners to achieve a greater intensity of labor per hour than was otherwise possible in an agrarian setting.

While equation (2) adds to the weight of evidence regarding the key role of the gang system, it is obviously not the final word.

This regression involves the assumption that the cotton South can be described by a homothetic production function, which may be inappropriate.[19] Moreover, the measure of land input employed in both equations (1) and (2) is uncorrected for the locational component of land values, a correction which is much more costly to perform on individual farms than in the aggregate for an entire size class. As our discussion of this correction indicated, the inclusion of the locational component of land values as an input serves to bias the scale coefficient, in both equations (1) and (2), downward. The results of our efforts to deal with these and other questions will be reported at a later date.

III. Reply to Schaefer and Schmitz

We agree with Donald Schaefer and Mark Schmitz's conclusion that total factor productivity in southern agriculture was positively correlated with farm size, as well as their conclusion that the positive association between farm size and total factor productivity should not be attributed to crop mix. But we do not agree with the analysis that led them to these conclusions.

The analysis of Schaefer and Schmitz turns on two relationships. The first is the correlation between total factor productivity (G) and the number of slaves (which may be taken as a proxy for an index of inputs),[20] when output (Q) is held constant. The second is the correlation between G and Q, when the number of slaves is held constant. Using both a tabular presentation and partial correlation analysis, Schaefer and Schmitz found that the first correlation had a negative sign and the second a positive one. Hence, they concluded that there were economies of scale, but that scale should be measured by output rather than by inputs.

Let us suppose that southern farms were characterized by a Cobb-Douglas production function with constant returns to scale, as shown by equation (3):

$$Q=AL^{\alpha_L}T^{\alpha_T}K^{\alpha_K}=AZ; \tag{3}$$

then

$$G=\frac{Q}{L^{\alpha_L}T^{\alpha_T}K^{\alpha_K}}=\frac{Q}{Z} \tag{4}$$

where Z is, as indicated, a geometric average of the inputs, and G, equals A.

Although economies of scale are absent, the Schaefer-Schmitz procedure would nevertheless show that G and Z are negatively correlated when Q is held constant, and that G and Q are positively correlated when Z is held constant. The point is that $r_{GZ.Q}$ would have a negative sign, and $r_{GQ.Z}$ a positive one, regardless of whether there were constant, increasing, or decreasing returns to scale, since in each case the signs obtained from these partial correlations are a necessary consequence of the definition of G. In other words, no inference about the existence or nonexistence of economies of scale can be made from their procedure.

If there were increasing or decreasing returns to scale, then equations (3) and (4) would become

$$Q=AZ^{1+\sigma} \tag{5}$$

$$G=AZ^{\sigma} \tag{6}$$

where σ is the scale coefficient ($\sigma=\Sigma\alpha_i-1$) and can be negative, zero, or positive. It follows from (5) and (6) that one can test for the existence of a scale effect by regressing Q or G on Z. As we have already seen, σ is positive and statistically significant, which indicates that G increased with the level of inputs.

Moreover, the Schaefer-Schmitz procedure will disguise the presence of a scale effect even when it actually exists. This can be demonstrated by considering the third panel in their Table 2, which shows the

---

[19] One should not leap to the conclusion that it must be incorrect. If the effect of the gang system was to augment labor rather than to shift the elasticities of the production function, the increase of comparative advantage with size would be consistent with a homothetic function.

[20] A regression of $\hat{Z}$ on $\hat{L}$, where Z is a geometric index of the inputs, yielded an elasticity of 1.08 ($t=178$) and an R^2 of 0.88, with $N=4299$.

TABLE 6—THE CLASS MEANS OF THE INDEX OF TOTAL FACTOR PRODUCTIVITY (G) IN THE SCHAEFER-SCHMITZ SUBSAMPLE OF FARMS AT OUTPUT LEVEL 3 ($\$1,500 < Q < \$5,000$) AND IN THE COMPLETE SAMPLE OF FARMS

Size Class as Measured by Number of Slaves (1)	Mean Values of G in the Complete Sample (2)	Mean Values of G in the Subsample with Output between \$1,500 and \$5,000 (3)	Percentage Deviation of the Subsample Mean from the Sample Mean (4)	Location of the Subsample Mean in the Complete Within-Class Distribution of G (5)
0	100.0	320.5	+220.5	bottom of 10th decile
1–15	107.7	154.8	+43.7	top of 6th decile
16–50	144.7	124.8	−13.8	bottom of 3rd decile
51+	133.5	54.4	−59.3	bottom of 1st decile

Source: Parker-Gallman sample.

relationship between size class (as measured by the number of slaves) and total factor productivity, within the subset of farms having an output level of at least \$1,500 but less than \$5,000. The effect of this procedure, as is shown by Table 6, is to force a comparison between a productivity index of free farms belonging to the tenth decile of the distribution of free farms and a productivity index of large slave farms belonging to the first decile of the distribution of large slave farms. In other words, by holding output constant, they introduced a bias into the comparison of G by input size class that is quite similar to the bias introduced by Wright when he held the cotton share constant. But in the Schaefer-Schmitz case the bias is more extreme. Consequently, they could not compare free and large slave farms in their lowest output level, since no large slave farm was so inefficient as to produce less than \$500 of output. Nor could they compare free and large slave farms in their highest output level, since only 2 out of 2,081 free farms were so efficient as to produce more than \$5,000 of output.

It should be noted that G and Q can be positively correlated under conditions of constant returns to scale, purely because of the role of the stochastic term. Consider the case where all farms not only have the production function shown by equation (3) but: (a) all farms have identical inputs; and (b) output is subject to stochastic as well as systematic factors. Then equations (3) and (4) become

$$Q = AZe^{\mu} \tag{7}$$

$$G = Ae^{\mu} \tag{8}$$

where μ is assumed to be normally distributed with mean zero. Under these conditions the correlation coefficient between G and Q will be equal to one, since correlating G with Q reduces to correlating μ with itself. On the other hand, there will be a zero correlation between G and Z. This example demonstrates that scale should be measured by inputs rather than by output. It also explains why Schaefer and Schmitz obtained a positive correlation between G and Q, even when the cotton share was held constant. While holding the cotton share constant served to control for the scale of measured inputs, it did not remove the positive correlation between G and Q due to μ.

IV. Concluding Comments

The principal function of the gang system was to increase the intensity of work per hour. The gang played a role comparable to the factory system or, at a later date, the assembly line, in regulating the pace of labor. It was, in other words, an early device for labor speedup. While the gang system provided a powerful mechanism for controlling the intensity of labor, slaveowners varied in their ability to exploit it. Various combinations of force and material incentives were used to elicit intense labor, but these were constrained by slave attitudes and desires. Although not to the same degree as employers of free labor, masters had to work out certain accommodations to the wishes of slaves.

Of the many unresolved issues raised by the debate, perhaps the most intriguing and difficult is the meaning of "harder work" or "more intense" labor. These terms are often loosely used, as though their meaning were perfectly obvious. Among the several possible definitions, we believe that two are most relevant. The first defines intensity of labor by caloric requirement. Thus one person (of a standard weight and height) would be said to work more intensely than another if his (or her) caloric requirement per hour of labor was greater. But one person could work more intensely than another even if both required the same number of calories. Such a situation would exist if the amount of motion of the two workers was identical, but one "wasted" less motion than the other. The slave case probably involved both kinds of intense labor—that is, both labor requiring more calories per man-hour and labor with less wasted motion. Labor which eliminates wasted motion may result in psychic fatigue and alienation, and so be more obnoxious than labor which permits wasted motion. The unremiting, machine-like quality of gang laborers repelled mid-nineteenth-century observers who valued traditional agrarian ways. And trade unions today frequently resist the introduction of practices aimed at eliminating wasted motion, even when workers are compensated by somewhat higher wages.

Wright raised another unresolved issue in the concluding section of his comment. Would the South have sustained its 1840–60 rate of growth in per capita income if slavery had continued? Wright contends that it would not have because the rate of growth in the British demand for cotton "had largely played out by 1860" (1979, p. 225). While we pointed out that the demand for cotton grew less rapidly after 1860 than before (see our 1971a paper, pp. 316–19, 330–31), the reduced growth rate of demand was still more rapid than the rate of growth of the South's agricultural labor force.[21] Furthermore, the substantial rise in the United States' production of cotton between 1875 and 1890 could not have occurred without either a substantial shift of resources from other activities into cotton, or an increase in the productivity of the cotton sector of agriculture, or some combination of both. Wright's counterfactual question regarding the course of the southern economy after 1860 cannot be settled merely by reference to the reduced rate of increase in cotton demand. The answer turns on the rate of productivity growth, not only in the various sectors of southern agriculture, but in nonagriculture as well, and on the capacity of the post-Civil War economy to reallocate resources in accord with shifting requirements. The past few years have produced several books and dissertations as well as a spate of shorter works which address these hitherto neglected issues. It would be premature to attempt to resolve Wright's interesting counterfactual question before the current burst of research into the economy of the postbellum South is completed and a reasonable picture of the actual course of events has emerged.

Opposition to the application of scientific methods to the study of history has often rested on the belief that these methods will divert attention from the great moral issues of history. This fear has not materialized in the study of slavery. Quite the contrary, cliometric research has served to emphasize the deeply moral nature of the antislavery crusade. Slavery was bad not because it was unprofitable and inefficient, or because it failed to deliver a high rate of economic growth, but because, in a world increasingly dedicated to personal liberty, it was the most malignant remnant of the past and the greatest obstacle to the continued expansion of the realm of individual choice. It was on this premise that the dissenters launched the antislavery crusade, and it remained their central premise to the end. If certain antislavery crusaders latched on to economic arguments, it was not because morality had given way to economic self-interest but because these crusaders thought that such arguments would win new adherents to their cause.

---

[21] The *world* demand for cotton continued to expand at the pre-Civil War rate until 1870. It stagnated during the depression decade of the 1870's and then resumed its upward course at about one-half the pre-Civil War rate. It should be noted that in comparing 1860 with 1877–78, Wright significantly underestimates the rate of growth in the demand for *U.S.* cotton, since 1860 was at a peak of a boom and 1878 was at the trough of the long recession that began in 1873. He also focuses on either British demand or the demand for *U.S.* cotton when it is total world demand that is relevant. As Lars Sandberg points out, the decline in British demand was offset by a rise in demand elsewhere in the world. The demand for *U.S.* cotton rose less rapidly between 1860 and 1890 than did total world demand, because the cotton famine of the Civil War induced other producers of cotton into the market. While the United States made rapid strides toward regaining its share, it had not fully closed the gap by the end of the 1880's. According to Michael Mulhall (p. 156), the United States counted for 78.1 percent of the world cotton production during 1851–1860. During 1881–87, its share was 72.0 percent.

REFERENCES

Arthur H. Cole, *Wholesale Commodity Prices in the United States, 1700–1861, Statistical Supplement,* Cambridge, Mass. 1938.

A. H. Conrad and J. R. Meyer, "The Economics of Slavery in the Ante Bellum South," *J. Polit. Econ.*, Apr. 1958, *66*, 95–130.

P. A. David and P. Temin, "Slavery: The Progressive Institution?," *J. Econ. Hist.*, Sept. 1974, *34*, 739–83.

______ **and** ______, "Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Comment," *Amer. Econ. Rev.*, Mar. 1979, *69*, 213–18.

Ezekiel J. Donnell, *Chronological and Statistical History of Cotton,* New York 1872.

R. A. Easterlin, "Regional Income Trends, 1840–1950," in Seymour Harris, ed., *American Economic History,* New York 1961.

Robert W. Fogel and Stanley L. Engerman, (1971a) "The Economics of Slavery," in their *The Reinterpretation of American Economic History,* New York 1971.

______ **and** ______, (1971b) "The Relative Efficiency of Slavery: A Comparison of Northern and Southern Agriculture in 1860," *Explor. Econ. Hist.*, Spring 1971, *8*, 353–67.

______ **and** ______, *Time on the Cross,* Vols. I, II, Boston 1974.

______ **and** ______, "Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South," *Amer. Econ. Rev.*, June 1977, *67*, 275–96.

______ **et al.,** *Without Consent or Contract: Technical Papers on Slavery,* New York 1981.

F. Linden, "Economic Democracy in the Slave South: An Appraisal of Some Recent Views," *J. Negro Hist.*, Apr. 1946, *31*, 140–89.

J. Metzer, "Rational Management, Modern Business Practices, and Economies of Scale in Ante-Bellum Southern Plantations," *Explor. Econ. Hist.*, Apr. 1975, *12*, 123–50.

Michael G. Mulhall, *The Dictionary of Statistics,* London 1899.

J. F. Olson, "Clock-Time vs. Real-Time: A Comparison of the Lengths of the Northern and Southern Agricultural Work-Years," mimeo., Univ. Connecticut 1976.

William N. Parker and Robert E. Gallman, "The 1860 Cotton Sample" (LCPSR 7419), in *Guide to Resources and Services, 1977-1978,* Univ. Michigan.

R. R. Russel, "The Effects of Slavery Upon Nonslaveholders in the Ante-Bellum South," *Agr. Hist.*, Apr. 1941, *15*, 112–26.

Lars G. Sandberg, *Lancashire in Decline: A Study in Entrepreneurship, Technology, and International Trade,* Columbus 1974.

D. F. Schaefer and M. D. Schmitz, "The Relative Efficiency of Slave Agriculture: A Comment," *Amer. Econ. Rev.*, Mar. 1979, *69*, 208–12.

George R. Taylor, *The Transportation Revolution, 1815–1860,* New York 1951.

G. Wright, "The Economics of Cotton in the Antebellum South," unpublished doctoral dissertation, Yale Univ. 1969.

______, "Slavery and the Cotton Boom," *Explor. Econ. Hist.*, Oct. 1975, *12*, 439–51.

______, "Prosperity, Progress and American Slavery," in Paul A. David et al., eds., *Reckoning With Slavery,* New York 1976.

______, "The Efficiency of Slavery: Another Interpretation," *Amer. Econ. Rev.*, Mar. 1979, *69*, 219–26.

American Cotton Planter, N.S. Vol. III, Montgomery, Sept. 1859.

U.S. Bureau of the Census, *Preliminary Report of the Eighth Census, 1860,* Washington 1862.

______, *Eleventh Census of the United States: 1890; Report on the Statistics of Agriculture in the United States,* Washington 1895.

On the Monopolistic Provision of Excludable Public Goods

By DAGOBERT L. BRITO AND WILLIAM H. OAKLAND*

The private market provision of excludable public goods is the subject of considerable controversy in economics.[1] Different analysts have produced results which range over the entire spectrum of possibilities—from extreme underprovision to complete overprovision. It is not our intention to review or to join in this voluminous debate, but rather to examine the question within a different and, in our view, more realistic context. Specifically, we examine the properties of private market provision under conditions of monopoly. By way of contrast, the existing literature has focused almost exclusively upon competitive or quasi-competitive provision.[2]

By way of definition, an excludable public good is a pure public good for which exclusion by means of price is costless. Since it is a pure public good, each consumer *could* consume the total quantity produced. Because exclusion is costless, however, there is no *need* for this to be the case. Hence, if q^i is the consumption of the ith consumer, $i=1,\ldots,N$, and Q is the total quantity produced, we have

$$q^i \leqslant Q \qquad i=1,\ldots,N$$

with the strict inequality holding whenever exclusion is exercised.

The emphasis upon competitive provision in the literature is somewhat surprising in view of the paucity of excludable public goods which are or could be provided within such a framework. The most convincing example to date is the television tape which, because of the absence of scale economies, is amenable to decentralized provision. Even here, however, one might argue that the relevant public good is the *transmission* of the tapes, an activity characterized by decreasing costs and formidable barriers to entry.

There is no such shortage of examples of candidates for monopoly provision. Highways, airports, waterways, parks, zoos, and museums are some services which come readily to mind. As long as the facility remains uncongested, each example meets the criteria for an excludable pure public good. Moreover, each enjoys scale economies in its production as well as a uniqueness arising because of locational considerations. If left in the private sector, such services would undoubtedly be characterized by monopoly elements. Thus, besides being of pure theoretical interest, the analysis of the monopoly provision of public goods carries considerable policy interest.

In what follows, a monopoly firm's behavior is examined under two different pricing arrangements. In the first, the firm is constrained to choose a single uniform price at which it will serve all of its customers. In the second, the firm can vary price with sales but must make the same price offers to each of its customers. Essentially this involves the selection of a non-linear outlay function for the firm's services. This function is then compared with that which would prevail if the industry were organized competitively.

Our principal findings are threefold: 1) Whether or not it is constrained to a single uniform price, a monopoly will produce a suboptimal level of public goods and inefficiently exclude some consumers from part of the output actually produced. 2) In general, a monopolist will find it profitable to vary price with the level of service provided to a particular customer. In a wide range of cases marginal price will be declining with respect to units of service provided. 3) If prices vary with units of service the monopoly price schedule will lie above the corresponding competitive price schedule and the monopoly output will fall short of the competitive level.

---

*Professors of economics, Tulane University, and senior fellows, Murphy Institute for the Study of Political Economy. This research was funded in part by a grant from the National Science Foundation. We wish to acknowledge useful comments by Michael Intriligator and the late Michael Visscher.

[1] See for example, Earl Thompson (1968, 1969), Bruce Owen, Harold Demsetz, Robert Ekelund and Joe Hulett, and Oakland.

[2] The only published treatment of the monopoly case has been by Thompson (1968). However, the assumptions employed depart considerably from those adopted here. For unpublished work, see Peter Tomlinson and Geoffrey Brennan, John Head, and Cliff Walsh.

I. Provision with a Single Price

Some of the most interesting properties of monopoly provision can be gleaned from a simple model in which the firm chooses a uniform price to be charged to all customers.[3] Such a single price could be the result of government restrictions, high costs of obtaining the necessary information for price discrimination, or because of the ability of customers to retrade. In any event, this case corresponds most closely to the analysis which is generally used to study the monopolistic provision of private goods and services.

To keep matters simple, we assume that the firm produces the public good under conditions of constant cost. Hence, total costs (C) equal unit costs (c) times total output(Q):

$$C=cQ \tag{1}$$

On the demand side, an individual's demand for the service is given by

$$d^i=D_i(p) \qquad i=1,\dots,N \tag{2}$$

It is customary in private good analysis to sum the demand curves (2) to obtain market demand. On the other hand, market demand for public goods is usually obtained by inverting (2) and summing. In the present case, however, neither procedure is appropriate. To sum demands "horizontally" is inconsistent with the public good property of the service. To sum "vertically" violates the condition that consumers be charged a common price. Aggregation of individual demand curves in the present case is much more complicated.[4]

The problem confronting the firm is to maximize

$$p\sum_{i=1}^{n} q^i-cQ \tag{3}$$

subject to

$$D_i(p)-q^i\geqslant 0 \tag{4}$$

$$Q-q^i\geqslant 0 \tag{5}$$

where q^i is the amount of service extended to the ith consumer. By (4), q^i is constrained to be no greater than an individual's demand d^i, while (5) restricts q^i to be no greater than total output, Q. It is (5) that expresses the nonrival character of public goods while (4) embodies the property of costless exclusion.

Defining λ^i and μ^i to be the shadow prices of constraints (4) and (5), respectively, and assuming $p>0$, $Q>0$, the Kuhn-Tucker conditions are

$$\sum_i q^i+\sum\lambda^i D_i'(p)=0 \tag{6}$$

$$\sum_i \mu^i-c=0 \tag{7}$$

$$\lambda^i\left[D_i'(p)-q^i\right]=0 \tag{8}$$

$$\lambda^i\geqslant 0 \qquad i=1,\dots,N$$

$$D_i(p)-q^i\geqslant 0$$

$$\mu^i\left[Q-q^i\right]=0 \tag{9}$$

$$\mu^i\geqslant 0 \qquad i=1,\dots,N$$

$$Q-q^i\geqslant 0$$

$$p=\lambda^i+\mu^i \qquad i=1,\dots,N \tag{10}$$

---

[3] An extensive discussion of this case is in Tomlinson.

[4] Since such a construct is not necessary for the solution we will not take the time to develop it here. For such a derivation, see Tomlinson.

It can be shown that consumers can be classified into two mutually exclusive sets: those for whom $q^i<Q$ and hence $\mu^i=0$; those for whom $q^i<D_i(p)$ and hence $\lambda^i=0$.[5] We label these sets R and $\bar{R}$, respectively. If the sets were not mutually exclusive, we would have, for some i

$$q^i<D_i(p)<Q$$

But this is not consistent with profit maximization because the firm, without changing p or Q, could allow q^i to increase to $D_i(p)$, thus adding $p\cdot[D_i(p)-q^i]$ to its profits.

Thus, from (10) we have

$$p=\mu^i \quad \lambda^i=0 \quad i \text{ in } \bar{R} \quad (Q=q^i<D_i(p))$$

$$p=\lambda^i \quad \mu^i=0 \quad i \text{ in } R \quad (D_i(p)=q^i<Q)$$

Substituting into (6) and (7) yields

$$(6') \qquad \sum_i q^i+p\sum_R D_i'(p)=0$$

$$(7') \qquad n(\bar{R})\cdot p=c$$

where $n(\bar{R})$ is the number of individuals in the set $\bar{R}$. Equation (7′) is the marginal revenue equals marginal cost condition for units of capacity. Since only members of set $\bar{R}$ consume the entire capacity, Q, only they are relevant to the output expansion decision. The firm can sell an additional unit of output to all members of $\bar{R}$ simultaneously since each member is constrained by quantity, not by price. Hence, the additional revenue from producing an additional unit of capacity is $n(\bar{R})\cdot p$.[6] Notice that this output expansion rule has the familiar public good flavor of adding up consumer demand prices. As we shall see below, this property carries over to a much wider class of models.

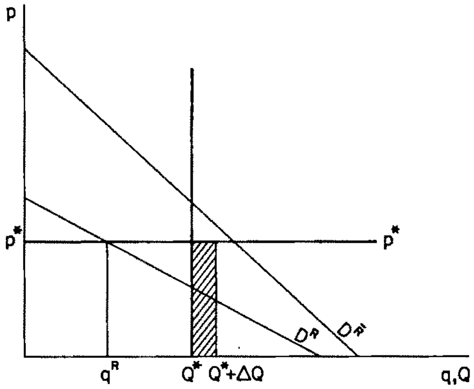


FIGURE 1

Turning to (6′), we see that the monopolist sets his price so as to maximize revenue for any given level of output, Q. The first term corresponds to the increased revenues obtained from inframarginal units, whereas the second term reflects the loss of sales due to the price increase. Since only those in set R are on their demand curves, the monopolist must only consider the loss in sales to this group.

A graphical interpretation of the solution is given by Figures 1 and 2. In Figure 1, the demand curves for representative members of groups R and $\bar{R}$ are shown. With price set at p^* and capacity set at Q^*, the consumption of the member of $\bar{R}$ will be at Q^*, below his/her preferred consumption. The consumption of the member of R is q^R, which is all that will be voluntarily consumed at price p^*. In deciding whether to expand capacity, say to $Q^*+\Delta Q$, the firm will only consider individuals in set $\bar{R}$, since at p^* only such individuals will consume additional output. If the shaded area shown in Figure 1, when multiplied by the number of members of $\bar{R}$, exceeds $c\Delta Q$, capacity expansion is warranted; otherwise it is not. Since $n(\bar{R})$ is a decreasing function of Q, a unique optimum capacity exists.

In Figure 2, we show the price-setting process. With capacity fixed at Q^* an increase in price of Δp will yield $Q^*\Delta p$ additional revenue from each member of $\bar{R}$ (they do not reduce their purchases because of excess demand at p^*). This increase in revenues must be compared with the loss of

---

[5] It is conceivable that for some i, $\mu^i\neq0$ and $\lambda^i\neq0$. However, if individual demands are unique and the number of individuals large, such consumers will be of measure zero; i.e., there are no mass points.

[6] That members of the set R are not relevant to the capacity decision owes to the fact that the firm always (whether capacity is expanded or not) has the option of selling them more. Of course, it must reduce p to do so.

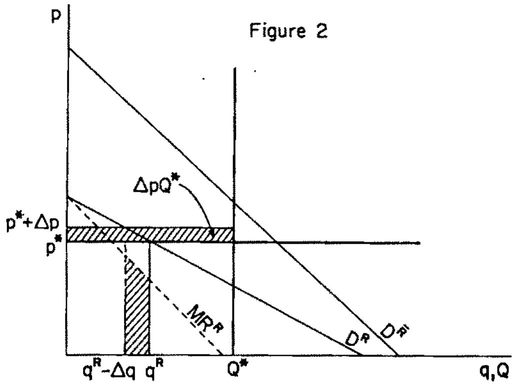


FIGURE 2

revenues from increasing price to the low-demand set.[7] The latter, for a typical individual, is given by the shaded area under the marginal revenue curve between q^R and $(q^R-\Delta q)$; that is, that loss of sales caused by the price increase. The price p^* will be optimal only if the total losses obtained by aggregating over the low-demand set cancel the gains derived from the high-demand set.

Let us turn to the welfare implications of this solution. First, the solution violates efficiency because those in set R are denied access to the complete bundle of services, Q, which are produced. Second, the solution is inefficient because, at the margin, the sum of individual marginal benefits, as measured by demand price, exceeds the marginal cost of producing an additional unit of public good. The very fact that nonmarginal members of set $\bar{R}$ have excess demand means that they would be willing to pay more than p^* for an additional unit. Yet, because the same price must be charged to all members of $\bar{R}$, this excess demand cannot be reflected in additional output.

The results of this model are quite similar to those reached by Oakland for the competitive case. However, a comparison of the results of the cases is best deferred until later in the paper when the behavioral assumptions of the two approaches are more consistent.

Three principal conclusions of the uniform price model can be drawn:

1) Consumers will be divided into two classes: those which are rationed by price and those rationed by output.

2) Only those consumers who are rationed by output will be relevant to the output expansion decision, and their demands will be reflected by price rather than marginal revenue.

3) The solution will not be Pareto efficient. The allocation of existing output will be inefficient because some consumers are excluded from consuming the total amount of good available. The level of output will be suboptimal because the profits of the firm fail to reflect the excess demand price premiums of some consumers.

That these conclusions are quite robust to the assumptions of the model will be evident below.

II. A Variable Price System

Thus far the analysis has proceeded under the assumption that the firm will charge a uniform price. We now relax this condition and assume that the firm is able to vary price with the quantity purchased. As before, however, we continue to assume that the firm does not discriminate *among* individuals. That is, the same price schedule must be available to all consumers.

The problem confronting the firm is to choose the revenue schedule $R(q)$ which will maximize

$$\sum_{i=1}^{N} R(q^i)-cQ$$

subject to

$$q^i \leqslant Q \tag{11}$$

and individuals' demand behavior. Individuals are assumed to maximize $u^i(x^i,q^i)$ subject to

$$x^i+R(q^i) \leqslant I^i \tag{12}$$

where x^i is a composite private good whose

---

[7] To be at a maximum, it is clear that those in the low-demand set must be in the elastic portion of their demand schedules (in the aggregate).

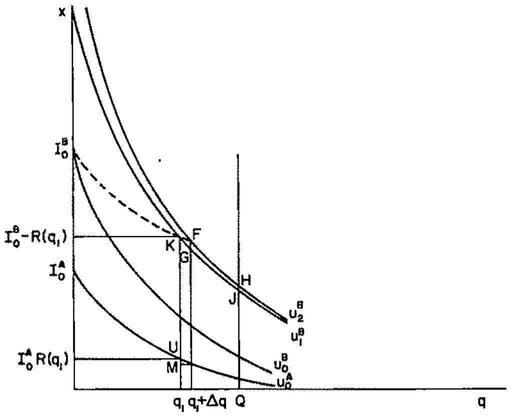


FIGURE 3

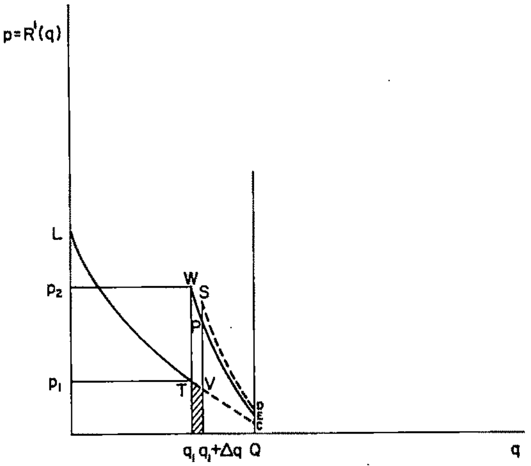


FIGURE 4

price is assumed to be unity, u^i is the individual's utility function, and I^i is his/her income. The solution to the individual is given by (12) and

$$(13) \qquad u_q^i/u_x^i=R'(q^i)$$

where u_q^i and u_x^i are the marginal utilities of public and private goods, respectively. The constraints on the firm are thus equations (11)–(13).

A graphical representation of the key elements of the problem is given in Figures 3 and 4. Let us assume initially that all individuals have identical tastes and incomes. In Figure 3, the curve labelled U_0^A represents the level of satisfaction which would be achieved with zero consumption of the public good. For this situation the monopolist would set $R(q)$ such that the consumer's budget restraint would coincide with the U_0^A curve. This implies the marginal price schedule, $R'(q)$, of LC in Figure 4. Because the consumer's marginal rate of substitution is nonincreasing everywhere on U_0^A, the $R'(q)$ schedule is also nonincreasing.

Next let us assume that individuals fall within one of two taste-income groups, A and B. The curve U_0^B in Figure 3 shows the zero public good utility level for those in B, while U_0^A serves that role for A. Suppose initially that the monopolist sets $R(q)$ such that those with income I_0^A consume q_1 and those with I_0^B choose to consume total capacity Q. This could be accomplished by an $R(q)$ schedule which makes the budget line of A coincident with U_0^A up to q_1 and coincident with U_1^B in the range q_1 to Q, where U_1^B is the utility members of B would attain if they purchased q_1 at the same terms as members of A. That such a policy dominates any other for the consumption bundles q_1 and Q can be seen from Figure 3. Clearly, the policy of running type B individuals down their indifference curve beyond q_1 is optimal because this will have no effect on A's choices. Suppose that q_1 were offered to A on more favorable terms so that they enjoyed some consumer surplus. This means that the budget line confronting B would be shifted upwards at consumption q_1—above the point K (where the curve $I_0^BK=I_0^AU$). This implies a utility level for B at q_1 which is greater than U_1^B, say U_2^B. But this means that at Q members of B will consume more private good. Since both A and B enjoy an increase in private good consumption, they must be spending less on the public good.

The marginal price schedule corresponding to the budget line I_0^BKJ is given by $LTWE$ in Figure 4. Notice that the price schedule jumps upward at q_1. This is due to the assumption that the income elasticity for the public good is positive; hence, the slope of B's indifference curves at q_1 exceeds that of A (on U_0^A at q_1). Beyond the discontinuity, the marginal price schedule falls as before along WE.

To determine whether q_1 is the profit-maximizing consumption for members of A given capacity Q, we compare the costs and benefits of extending consumption by Δq, as in Figure 4. Members of A will pay $q_1TV(q_1+\Delta q)$ for the additional consumption. In order to induce members of A to consume the Δq units, the monopolist must reduce the marginal price on these units. This results in a loss of $TWPV$ from members of B. However, the latter is partially offset by an increase in willingness to pay for the remaining units, from the locus PE to SD; the monopolist thus earns an additional $PSDE$ on those units per member of B. If n^A and n^B correspond to the number of individuals in groups A and B, respectively, the change in the monopolist's total profit (Π) is

$$\Delta\Pi = n_A \cdot q_1TV(q_1+\Delta q) - n_B[TWPV - PSDE]$$

For maximum profits this expression must be zero. This expression can also be interpreted in terms of Figure 3. The increased sales to A are given by n_AUM, which for small Δq is $n_Ap_1\Delta q$ where p_1 is the marginal price faced by A at q_1. The loss in revenue from B is given by n_BJH, the increase in private good consumption caused by the reduced marginal price on the Δq units. Because q is normal, JH must be less than FG. But we have $FG = p_2\Delta q$; hence the change in the monopolist's revenues is

$$\Delta\Pi = (n_Ap_1 - \gamma n_Bp_2)\Delta q$$

with $\gamma < 1$. The condition for maximum profits is

$$\frac{p_1}{p_2} = \gamma\frac{n_B}{n_A}$$

Unfortunately, this solution is not in closed form since γ itself depends upon n_A and n_B. Hence, it is not possible to ascertain from the graphical analysis so far whether an increase in n_A/n_B will raise or lower p_1/p_2.

The preceding analysis makes it clear that the capacity expansion decision has the properties of the uniform price model. Figure 3 indicates that changes in Q only involve the marginal valuations of those in the high-demand class. In effect, the choice of capacity is independent of the inframarginal prices charged for consumption. Hence, as long as it is profitable to exclude members of A from some part of total output, they are irrelevant for the choice of capacity.

The introduction of additional income classes into the model will not change the essential characteristics of the solution described above. Discrete differences in income will always cause the discrete jumps in marginal price such as TW in Figure 4. As income differences narrow, however, the size of the jump will also narrow until they disappear altogether with a continuous income distribution. Unfortunately, however, the analysis of the type used in Figures 3 and 4 is incapable of predicting whether the marginal price schedule, $R'(q)$ will be increasing or decreasing. Indeed, even in the two-class case it is not possible to infer whether points such as L lie above points like W. Since, with a continuous income distribution, the optimum price schedule would be the locus of points like L and W, it seems clear that in general nothing definite can be said about the slope of $R'(q)$. Fortunately, however, other tools of analysis enable us to obtain specific results under a broad set of conditions. It is to this task we now turn.

A. *An Optimal Control Approach*

Considerable insight into the problem can be gained by casting the problem into the framework used by James Mirrlees to address the problem of optimal income taxation. In this approach, a criterion function is maximized subject to the constraints imposed by consumer market behavior. As before, we assume consumer preferences to be represented by a continuously differentiable strictly concave utility function $u(x,q)$. Although they have identical tastes, incomes of consumers vary continuously in the closed interval $[I^*, I^{**}]$ according to the density function $f(I)$.[8]

The maximand of the firm thus becomes

$$\int_{I^*}^{I^{**}} R[q(I)]f(I)dI - cQ$$

Since aggregate income Y can be taken to be fixed, the consumer's budget constraint can be used to write the maximand as

$$Y - \int_{I^*}^{I^{**}} x(I)f(I)dI - cQ$$

Hence, the maximization of the monopolist's revenue is equivalent to the minimization of aggregate consumption of the private good.

To obtain the constraints imposed by consumer market behavior, we totally differentiate the utility function to obtain

$$(14)\qquad \frac{du}{dI} = \left[-u_x R'(q) + u_q\right]\frac{dq}{dI} + u_x$$

which, using (13), reduces to

$$(15)\qquad \frac{du}{dI} = u_x(x(I), q(I))$$

Since this constraint on individual behavior only contains information from the first-order conditions of the consumer problem, it is conceivable that the firm's solution to its optimal pricing problem may not be consistent with consumer's choices. Such would be the case if the public good is "normal" (i.e., has positive income effects) and if the firm's solution is such that $dq/dI < 0$ in some interval.

To see this, suppose that in the interval $[I_1, I_2]$ the firm would like to have $dq/dI < 0$. This means that there is some income level $I_0 < I_1$ for which $q(I_0) = q(I_2)$. Since q is normal and since the same price schedule must confront consumers, the price schedule must have a corner at $q(I_0)$. If it does, however, the consumer at income level I_1 will also choose $q(I_0)$, which is inconsistent with the firm's plans. Since the assumption of normality for the public good is reasonable, the firm's choices must be restricted such that $dq/dI \geqslant 0$. This will be done by defining $v = dq/dI$ and requiring $v \geqslant 0$. Whenever this constraint is binding, the optimal price schedule will have a corner point.

The problem confronting the firm is to maximize profit subject to (15) and $v \geqslant 0$. By inverting the utility function, $x = x(u,q)$, x can be explicitly eliminated from the problem. The Hamiltonian is

$$(16)\qquad H = -x(u,q)f(I) + \Phi u_x(x(u,q),q) + pv - \delta v$$

where Φ and p are the costate variables associated with du/dI and dq/dI, respectively, and δ is the shadow price associated with the constraint on v.

The necessary conditions associated with (16) are

$$(17)\qquad \frac{d\Phi}{dI} = (f(I) - \Phi u_{xx})\frac{1}{u_x}$$

$$(18)\qquad \frac{dp}{dI} = -\frac{u_q}{u_x}(f(I) - \Phi u_{xx}) - \Phi u_{xq}$$

$$(19)\qquad p - \delta \leqslant 0;\ v(p-\delta) = 0$$

In addition we have the transversality conditions

$$(20)\qquad \Phi(I^{**}) = 0$$

$$(21)\qquad p(I^{**}) = -c$$

and the initial condition,

$$(22)\qquad u(I^*) \geqslant u(I^*, 0)$$

These boundary conditions have straightforward interpretations. The term $\Phi(I)$ represents the cost of increasing utility to the Ith income class. Since this cost reflects the additional amount of x which is given to higher income classes as the result of the increased utility of the Ith class, it obviously vanishes at the top. Similarly, $-p$ corresponds to the benefits of increasing public good consumption for the Ith class. Since the public good is always fully consumed by the I^{**}th class (otherwise production could be reduced), the only way of increasing its consumption is to produce additional units. Equation (21) simply states that marginal benefits equals marginal cost. Finally, equation (22) states that the utility of the lowest income class can never be less than that which they could achieve by not consuming the public good.

---

[8] Alternatively, it could be assumed that people have identical incomes but varying tastes. Such an approach would yield conclusions similar to those of this paper. It would be desirable to allow individuals to vary in both dimensions, but so far such an approach has not proved tractable.

Let I_1 be the lowest income class that consumes the total quantity of public good produced. It can be shown that (proofs are given in the Appendix):

PROPOSITION 1: *To maximize profits the firm must set the price of the last unit of output equal to marginal cost of production divided by the number of individuals consuming that unit. That is,*

$$R'(Q)=\frac{c}{N(I_1,I^{**})} \tag{23}$$

where $N(i,j)$ is the number of individuals with incomes between i and j.

PROPOSITION 2: *The monopolist will choose a level of output which is suboptimal in the Pareto sense. That is,*

$$\int_{I_1}^{I^{**}}\frac{u_q(s)}{u_x(s)}f(s)\,ds>c \tag{24}$$

Expression (23) is identical to the condition (7′) derived in the uniform price case. It implies that, at the optimum, marginal cost equals total revenue from selling the last unit of q.

Proposition 2 implies that consumers would be capable of bribing the firm to produce additional units of output. They cannot do so, however, because this would require personalized prices. The firm, by assumption, must charge the same price schedule to all customers.

PROPOSITION 3: *If $f(I)$ is continuous and it is desirable to exclude some individuals entirely from the consumption of the public good, the profit-maximizing price schedule does not contain a fixed charge. That is,*

$$\lim_{q\to 0}R(q)=0$$

Let the lowest income class consuming the public good be $I_0>I^*$. The existence of a fixed charge means that consumers with I_0 must be given a discrete bundle of q to compensate for the charge, since the individual has the choice of not consuming the public good. Hence $q(I)$ must be discontinuous at I_0. As shown in the Appendix this is inconsistent because of the continuity of the Hamiltonian in u, I, and q.

If all consumers consume some positive q, or if $f(I)$ is discontinuous, a fixed charge may be consistent with profit maximization. However, the monopolist could do just as well with a variable price schedule which moves individuals with I^* along the indifference curve $u(I^*,0)$ as was done in constructing Figure 4. In this sense, therefore, a fixed charge is never essential for profit maximization.

B. *Separable Utility Function*

It is difficult to determine other properties of the profit-maximizing price schedule without placing additional restrictions on the utility and density functions. In this section we consider the special case where the utility function is separable. Since cross derivatives are zero, (17) and (18) can be combined to yield

$$d\Phi/dI=-u_q\,dp/dI \tag{25}$$

Hence, $dp/dI=0$ implies that $d\Phi/dI=0$ and that (17) or (18) can be solved to obtain x as a function of I alone. Now $dp/dI=0$ whenever $\delta=0$ (i.e., the $dq/dI\geq 0$ constraint is not binding). Because the Hamiltonian is linear in v, $p=0$ whenever $\delta=0$; otherwise, the Hamiltonian could be increased without bounds.

Thus, in any open interval such that $\delta = p = 0$, we have from (17)

$$u_{xx}(x) = f(I)/\Phi \tag{26}$$

Since Φ is constant we are able to solve for x in terms of I. This yields the following proposition.

PROPOSITION 4: *Consider the interval* $[I_0, I_1]$ *such that* $q(I) = 0$, $I < I_0$, *and* $q(I) = Q, I > I_1$. *If the utility function is separable and the density is differentiable and nondecreasing, consumption of the private good will be strictly nonincreasing in the interval and marginal price,* $R'(q)$, *will be everywhere falling.*

How can we explain this tendency for the marginal price to fall? Recall that at points where $\delta = 0$, the firm is free to choose q so as to maximize the Hamiltonian. Since the latter, in the separable case, implies $d\Phi/dI = 0$, the indirect cost of decreasing the price of q to obtain more revenue from a class is confined to those individuals who consume capacity, Q. Hence, the cost is the same for all classes for whom $q < Q$. The direct benefits of reducing price to a class, on the other hand, is proportional to the numbers of members in the class, $f(I)$; hence, this consideration differentiates among classes when deciding whether to reduce its marginal price. In the case where $f(I)$ is constant, x will be constant. But with increasing q this can only be achieved by falling marginal prices. If $f'(I) > 0$, $dx/dI < 0$, requiring an even more sharply falling marginal price schedule.

If $x(I)$ is increasing (as would be the case if $f'(I) < 0$), the behavior of marginal prices may be ambiguous. Necessary and sufficient conditions for the marginal price schedule to be increasing are derived in the Appendix. These conditions are

$$1 > \frac{\dfrac{d \ln f(I)}{dI}}{\dfrac{d \ln u_{xx}}{dx}} > \frac{u_{qq} u_x^2}{u_{qq} u_x^2 + u_q^2 u_{xx}} \tag{27}$$

If the right-hand side inequality is not satisfied, marginal price will decrease as before. If the left-hand side is violated, the marginal price schedule will be discontinuous. At such points the firm would prefer that $dq/dI < 0$, but is restrained by individual choice from doing so. It is easy to see that at such points $\bar{q}$,

$$\lim_{q \to \bar{q}^+} R'(q) > \lim_{q \to \bar{q}^-} R'(q)$$

That is, marginal price is lower on the left-hand side of the discontinuity than on the right. Otherwise, $dq/dI > 0$ because of the convexity of the individual's indifference curves. In this sense, then, we can state that if the right-hand side of (27) is satisfied, marginal price will be increasing.

While we have restricted attention to cases where $f(I)$ is monotonic, it is possible to use these results to infer the shape of the price schedule when individuals are distributed in other ways. One pattern of obvious importance is where $f'(I) > 0$ to begin with but with $f'(I) < 0$ at higher income levels. In all such cases we are able to show that if individuals in the rising portion of the income distribution are allowed to consume at all, the marginal price schedule will be initially falling. Whether or not prices eventually increase, however, depends upon how fast density falls off relative to the decrease in u_{xx}.

III. Comparison With Competitive Results

A central assumption of this paper is that the firm must offer the same price schedule to all customers. This assumption is quite similar to that employed by Oakland in analyzing the problem under the assumption of free entry. A major result of his 1974 study was that consumers would face a rising price schedule for the public good. The price for successive units of consumption rises because the degree of utilization of successive units declines. Free entry and the requirement that profits be nonnegative on each unit of public good imply that the price of the marginal unit consumed by an individual satisfies

$$R'(q) = c/n \tag{28}$$

where c is the cost of production and n is the number of individuals with total consumption at least as great as that of the particular individual. Since n will be a noncreasing function of q, $R'(q)$ must also be nonincreasing. These price differentials among the various units of consumption cannot be eliminated by arbitrage because individuals will naturally consume those units with lower prices first. Hence, an individual whose marginal consumption price is high cannot trade with individuals whose marginal price is low.

In the context of this paper, (28) is equivalent to

$$(28') \quad R_c'(q(I)) = \frac{c}{N(I, I^{**})} \qquad I_0^c \leq I \leq I_1^c$$

where I_0^c and I_1^c are the lowest income classes consuming any and all units of the public good, respectively, and I_0^c and I_1^c are determined by

$$(29) \quad I_0^c = \min_I \frac{c}{N(I, I^{**})} \geq \frac{u_q(I,0)}{u_x(I,0)}$$

$$(30) \quad I_1^c = \max_I \frac{c}{N(I, I^{**})} \leq \frac{u_q(q(I), x(I))}{u_x(q(I), x(I))}$$

Thus, those with $I < I_0^c$ are unwilling to pay the price that competitive firms require for the first unit of output, while those with $I > I_1^c$ are unwilling to pay the marginal cost of capacity expansion if charged a common price.

A central feature of the competitive solution is that each unit of public good yields revenue exactly equal to its cost of production. Alternatively, each unit is sold at the lowest price consistent with nonnegative profits. Moreover, since $N(I, I^{**})$ is nonincreasing and $q(I)$ nondecreasing, $R_c'(q)$ is a nondecreasing function of both q and I.

How do the monopoly and competitive solutions compare? In what follows, we shall establish

PROPOSITION 5: *The monopoly marginal price schedule lies somewhere above and nowhere below the competitive solution, that is,*

$$(31) \quad R_m'(q) \geq R_c'(q) \qquad 0 \leq q \leq Q$$

$$R_m'(q) > R_c'(q) \quad \textit{some } q \textit{ in } [0, Q]$$

where the c and m subscripts refer to competition and monopoly, respectively.

If the monopoly solution is nonincreasing, the proof is simple. The opposite slopes of the schedules together with (23) and (28′) directly imply the proposition. To prove the general case we first show that the monopoly solution can never lie below the competitive price schedule at the outset; i.e., $R_m'(0) \geq R_c'(0)$. Suppose the contrary were the case. By construction the competitive price is the lowest consistent with nonnegative profits for the first unit. Hence, the monopolist must suffer a loss on this unit. This is inconsistent with profit maximization because the firm can do better by producing one unit less and defining a new price schedule, $\bar{R}_m(q) = R_m(q+1) - R_m(1)$, for $q = 1, \ldots, Q-1$. Thus, for example, we have $\bar{R}_m'(0) = R_m'(1)$, etc. Because the demand for the public good is normal, *no* individual will reduce his/her total purchases by more than one unit when the new price schedule is substituted for the old.[9] It follows that the monopolist's revenues from the $Q-1$ remaining units are no lower than before—indeed, they will be higher. Since it avoids a loss on the original first unit, therefore, the new price schedule $\bar{R}_m'(q)$ is more profitable than $R_m'(q)$.

The implication of the preceding result is that if the monopoly price schedule is to lie below the competitive schedule in some

---

[9] To prove the first part of Proposition 5, note that $\bar{R}_m(q) - R_m(q+1) = -R_m(1) < 0$. Hence, for any q, private good consumption x is greater than before, say, $\bar{x}(q) > x(q+1)$ for all I. Because, of the normality of the demand for q and x, the absolute value of the slope of an individual's indifference curve through the point $(\bar{x}(q), q)$ is greater than that through $(x(q+1), q+1)$. Since the slope of the price schedule at the two points is equal (by construction) and the price schedule is tangent to the indifference curve at $(x(q+1), q+1)$, it follows that the individual will consume more than q under the price schedule $\bar{R}_m(q)$. Hence, consumption is reduced by less than one unit.

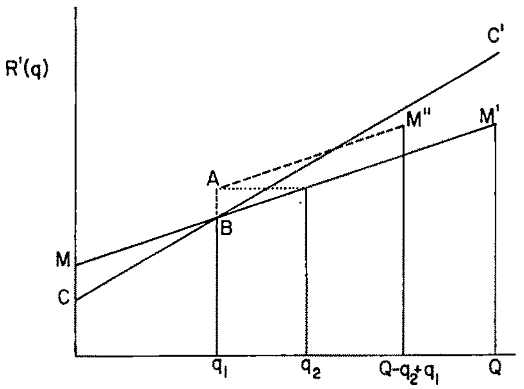


FIGURE 5

range, the former must intersect the latter from above at at least one point. This is shown in Figure 5, where for simplicity, linear marginal price schedules are shown. The monopoly schedule *MM'* intersects the competitive schedule *CC'* at q_1. Because of the normality of demand for q, the marginal income class at the point of intersection, q_1, must be lower under competition than under monopoly. Hence, the number of individuals consuming the q_1'th unit must be larger under competition. Since competitive firms break even on all units, the monopolist must suffer a loss on the q_1'th unit. This will be true of units $q>q_1$ such that $N(I_c(q), I^{**}) \geqslant N(I_m(q), I^{**})$; for not only will the monopolist be selling to fewer individuals but receiving a lower price as well. Thus, if the monopoly price schedule is profit maximizing, there must exist some $q_2>q_1$ such that the number of individuals consuming the last unit is equal under both market arrangements. Otherwise, it would be profitable for the monopolist to cease production at q_1. If such a q_2 exists, the monopolist can increase profits by reducing output by (q_2-q_1) units and setting the marginal prices for units q_1+h at the same levels as it did for q_2+h units before with $h>0$. This is shown by the curve *MBAM''* of Figure 5. Just as in the case of the first unit, consumers with incomes greater than $I_m(q_2)$ will decrease their purchases by no more than q_2-q_1 units. Of course, those with $I \leqslant I_m(q_1)$ will purchase just as before. Hence, the firm avoids the loss on the q_2-q_1 units and at the same time suffers no loss in revenues on its remaining units. Since this argument can be applied whenever the monopoly schedule intersects the competitive schedule from above, the proposition is proved in its weak form, i.e., $R_m'(q) \geqslant R_c'(q)$. That the strict inequality will hold for some q follows from the observation that the monopolist can always do better than employing the competitive price schedule.

Given Proposition 5 it is trivial to establish the next proposition.

PROPOSITION 6: *Monopoly output will always fall short of competitive output.*

This follows directly from the normality of demand for the public good. Elsewhere (see Oakland) competitive provision of excludable public goods was shown to lead to suboptimal production and underutilization of those goods actually produced. Proposition 6 states that underproduction is made even worse by monopoly provision. The problem of underutilization is also exacerbated because with the higher prices, monopoly will result in greater exclusion from the stock of goods in existence.

While the result that monopoly is less efficient than competition is not new in economic analysis, it is important that it be rigorously demonstrated for the case of excludable public goods. For the existing literature on the subject has provided many unorthodox conclusions. Indeed, using a different framework, Thompson (1968) argued just the opposite about the relative efficiency of monopoly and competition. The source of the difference in our efficiency conclusions stems from differences in assumptions about the firm's ability to discriminate.

In this paper, the firm can discriminate only among units of output, not among individuals per se. Thompson's firm, on the other hand, could discriminate in both dimensions. Since a perfectly discriminating monopolist can charge each customer his/her marginal valuation for each unit, it follows that the resulting allocation is economically efficient. However, the informational requirements for pursuing perfect price discrimination are so vast as to cast doubt on the practical significance of arguments based upon it. By way of contrast, our approach requires only that the monopolist have information about the income elasticity of demand for its product. Even if it totally lacks such information, however, our analysis of the uniform price policy suggests that the basic efficiency shortcomings of monopoly provision will remain.

IV. Concluding Remarks

Our primary objective in this paper was to explore the characteristics of monopoly provision of excludable public goods. One of our most important conclusions is the profitability of nonuniform price schedules. In particular, there are powerful forces which give rise to quantity discounts. Such a practice is not without empirical precedence; declining block structures are commonplace in the public utility arena. To date such practices have been explained by models of interpersonal price discrimination. Our analysis indicates that declining price structures can be explained without appealing to market segmentation.

In addition, the paper makes a methodological contribution. In particular, the technique developed here can be used with minor modification to explore pricing behavior for private goods. The assumption of a uniform price in imperfect markets can only be justified historically in terms of analytic convenience. However, with economists becoming increasingly familiar with the principles of optimal control, models which provide for non-linear pricing arrangements should be increasingly emphasized.

APPENDIX

The problem confronting the firm is

$$\max Y - \int_{I^*}^{I^{**}} x(u,q) f(I)\, dI - cQ$$

subject to $du/dI = u_x(x(u,q),q)$; $dq/dI = v \geqslant 0$; $q \leqslant Q$. The Hamiltonian for this problem is

$$\text{(A1)}\quad H = -x(u,q)f(I) + \Phi u_x(x(u,q),q) + pv - \delta v$$

where Φ and p are the costate variables associated with du/dI and dq/dI constraints, respectively, and δ is the shadow price associated with the nonnegativity constraint on v. The costate variables must satisfy the differential equations

$$\text{(A2)}\quad \frac{d\Phi}{dI} = -\frac{\partial H}{\partial u} = [f(I) - \Phi u_{xx}]\frac{1}{u_x}$$

$$\text{(A3)}\quad \frac{dp}{dI} = -\frac{\partial H}{\partial q} = -[f(I) - \Phi u_{xx}]\frac{u_q}{u_x} - \Phi u_{xq}$$

and the transversality conditions

$$\text{(A4)}\quad \Phi(I^{**}) = 0$$

$$\text{(A5)}\quad p(I^{**}) = -c$$

The first-order conditions are

$$\text{(A6)}\quad p - \delta \leqslant 0$$

$$\text{(A7)}\quad v[p - \delta] = 0$$

Note that the first-order conditions imply that if $p < 0$ then $\delta < 0$, and $v = 0$. Also if $\delta = 0$ in some closed interval $[I_1, I_2]$ then $p = 0$ in $[I_1, I_2]$ and $dp/dI = 0$ in (I_1, I_2). This implies that if the $v \geqslant 0$ constraint is not binding,

$$\text{(A8)}\quad [f(I) - \Phi u_{xx}]\frac{u_q}{u_x} - \Phi u_{xq} = 0$$

almost everywhere. This would be the first-order condition if q were a control variable.

Equation (A2) can be solved by introducing an integrating factor of the form

$$\eta(I,k) = \exp \int_I^k \frac{u_{xx}}{u_x} ds$$

The solution to the differential equation is

$$\Phi(I) = \frac{\Phi(I^*)}{\eta[I^*, I]} + \int_{I^*}^{I} \eta(I,s)\frac{f(s)}{u_x(s)} ds$$

where $\Phi(I^{**}) = 0$ implies that

$$\text{(A9)}\quad \Phi(I) = -\int_I^{I^{**}} \eta(I,s)\frac{f(s)}{u_x(s)} ds$$

and $\Phi(I)$ depends only on higher income classes. Furthermore since $\Phi(I^{**})=0$ and $\Phi(I)>0$ implies $d\Phi/dI>0$, it must be that $\Phi(I)<0$ for $I\leqslant I^{**}$.

Equation (A3) can also be integrated. If $dq/dI>0$, $p=0$; so consider the lowest income class, I_1, for which $q=Q$; for all $I\geqslant I_1$, $x(I)=x(I_1)+(I-I_1)$ and $q=Q$. This permits the integration by parts of the following expression which is obtained by combining (A2) and (A3).

$$\frac{dp}{dI}=-\left[\frac{d\Phi}{dI}u_q(x,Q)+\Phi u_{xq}(x,Q)\right]$$

$$\text{(A10)}\quad p(I)-p(I_1)=-\Phi(s)u_q(s)\Big|_{I_1}^{I}+\int_{I_1}^{I}\frac{d\Phi(s)}{ds}u_q(s)\,ds-\int_{I_1}^{I}\frac{d\Phi(s)}{ds}u_q(s)ds$$

The two integrals cancel and since $p(I^{**})=-c$, $p(I_1)=0$ and $\Phi(I^{**})=0$,

$$\text{(A11)}\qquad \Phi(I_1)u_q(I_1)=-c$$

For $I>I_1$

$$\text{(A12)}\quad \exp\int_{I_1}^{s}\frac{u_{xx}(m)}{u_x(m)}dm=\exp ln\frac{u_x(s)}{u_x(I_1)}=\frac{u_x(s)}{u_x(I_1)}$$

Combining equations (A9), (A11), and (A12) yields

$$\text{(A13)}\qquad \frac{u_q(I_1)}{u_x(I_1)}\int_{I_1}^{I^{**}}f(s)\,ds=c$$

This proves Proposition 1.

To prove Proposition 2, note that the normality of the public good requires that $u_q(I)/u_x(I)>u_q(I_1)/u_x(I_1)$ for all $I>I_1$. Thus

$$\text{(A14)}\quad \int_{I_1}^{I^{**}}\frac{u_q(s)}{u_x(s)}f(s)\,ds>\int_{I_1}^{I^{**}}\frac{u_q(I_1)}{u_x(I_1)}f(s)\,ds=c$$

To prove Proposition 3 suppose the contrary; the continuity of $u(x,q)$ and the existence of a fixed charge imply that $q^*(I_0)>0$ and $\partial H[I_0,q^*(I_0)]/\partial q=0$. The continuity of $f(I)$ implies that $\partial H/\partial q$ is a continuous function of I. This together with the concavity of H implies there exists $0<\hat{q}<q^*(I_0)$ such that $\partial H(I_0,\hat{q})/\partial q>0$. This implies there exists some ball $B_\epsilon(I_0,\hat{q})$ such that $\partial H(I,q)/\partial q>0$ for all I,q in $B_\epsilon(I_0,\hat{q})$. Consider some $\hat{I}<I_0$ in $B_\epsilon(I_0,\hat{q})$, $\partial H(I,\hat{q})/\partial q>0$ for all I in $(\hat{I},I_0)$. The condition that H is concave implies that $\partial H(I,0)/\partial q>0$, for I in $(\hat{I},I_0)$, which implies that $dp/dI<0$ in $(\hat{I},I_0)$. The condition that $q^*=0$ for I in $(\hat{I},I_0)$ implies $p(I)<0$ for I in $(\hat{I},I_0)$, so $dp/dI<0$ in $(\hat{I},I_0)$ implies $p(I_0)<0$. This implies that $q^*(I_0)=0$ which is contrary to assumption.

If the utility function is separable then equations (A2) and (A3) can be combined to yield

$$\text{(A15)}\qquad \frac{dp}{dI}=-u_q\frac{d\Phi}{dI}$$

Since $dp/dI=0$ almost everywhere if the $v\geqslant 0$ constraint is not binding, $\Phi^*(I)$ is a constant almost everywhere over the same interval.

In the interior of any interval where $dq^*/dI>0$, the problem can be treated as if q were the control variable; thus in the interior of such an interval it is necessary that the Hamiltonian satisfy first- and second-order conditions for a maximum. Thus,

$$\text{(A16)}\qquad [-f(I)+\Phi^*u_{xx}]\frac{\partial x}{\partial q}=0$$

and

$$\text{(A17)}\qquad \Phi^*u_{xxx}\left(\frac{\partial x}{\partial q}\right)^2<0$$

Since $\Phi^*<0$, (A17) can hold only if $u_{xxx}>0$. Note that $u_{xxx}>0$ implies a condition similar to decreasing absolute risk aversion in the uncertainty literature. If $u_{xxx}<0$, the firm's optimal strategy is to set a single, all-or-nothing change for total Q. Differentiating (A16) with respect to I yields

$$\text{(A18)}\qquad \frac{dx}{dI}=\frac{f'(I)}{\Phi^*u_{xxx}}$$

Since $u_{xxx} > 0$ and $\Phi^* < 0$, $f'(I) \geqslant 0$ implies $dx/dI \leqslant 0$. This proves the first part of Proposition 4. Individual maximization implies that $R'(q) = u_q/u_x$. Thus

$$R''(q)\frac{dq}{dI} = \frac{u_{qq}u_x\frac{dq}{dI} - u_q u_{xx}\frac{dx}{dI}}{(u_{xx})^2} \tag{A19}$$

Since $dq/dI > 0$ for an interior solution, and $dx/dI < 0$, $R''(q) < 0$. This proves the second part of Proposition 4.

If $f'(I) < 0$, (A18) shows that $x(I)$ will be increasing. Using the individual's budget constraint, we can reexpress (A19) as

(A20)

$$R''(q) = \frac{u_{qq}u_x^2 - \left[u_{qq}u_x^2 + u_q^2 u_{xx}\right]\frac{dx}{dI}}{u_q u_{xx}^2}$$

If $dq/dI > 0$ (i.e., $dx/dI < 1$), we have $R'' \lesseqgtr 0$ as

$$1 > \frac{dx}{dI} \gtreqless \frac{u_{qq}u_x^2}{u_{qq}u_x^2 + u_q^2 u_{xx}} \tag{A21}$$

Substituting from (A18) and (A16) yields inequality (27) in the text.

REFERENCES

G. Brennan, J. Head, and C. Walsh, "Market Provision of Public Goods: A Monopoly Version of the Oakland Model," *Finanzarchiv*, No. 3, 1979, *37*, 385–95.

H. Demsetz, "The Private Production of Public Goods," *J. Law Econ.*, Oct. 1970, *13*, 292–306.

R. Ekelund and J. Hulett, "Joint Supply, the Taussig-Pigou Controversy, and the Competitive Supply of Public Goods," *J. Law Econ.*, Oct. 1973, *16*, 369–87.

J. Mirrlees, "An Exploration in the Theory of Optimum Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, 175–208.

W. Oakland, "Public Goods, Perfect Competition and Underproduction," *J. Polit. Econ.*, Sept./Oct. 1974, *82*, 927–39.

B. M. Owen, "The Perfectly Competitive Production of Collective Goods: Comment," *Rev. Econ. Statist.*, Nov. 1969, *51*, 475–76.

E. A. Thompson, "The Perfectly Competitive Production of Public Goods," *Rev. Econ. Statist.*, Feb. 1968, *50*, 1–12; "A Reply," *Rev. Econ. Statist.*, Nov. 1969, *51*, 479–82.

P. Tomlinson, "Market Mechanisms and the Production of Public Goods," unpublished doctoral dissertation, Johns Hopkins Univ. 1973.

Taxation, Human Capital, and Uncertainty

By JONATHAN EATON AND HARVEY S. ROSEN*

The growth of tax rates on earned income has focused attention upon the impact of such taxes on work effort. The effect of taxes on hours of work has been the subject of both theoretical and econometric study.[1] The basic lesson of the theoretical literature is that the impact of taxation on hours of work is logically indeterminate because of the familiar conflict between income and substitution effects. Therefore, an enormous amount of econometric research has been done on this subject. Although there is a disconcertingly high variance in estimates of labor supply elasticities, it would be fair to say that the consensus is that the response in hours of work to changes in the net wage is very small for prime age male earners.[2] However, "work effort" and "hours of work" are not synonymous. Even if taxes have little impact upon quantity of hours worked, it is possible that they have an important effect on their quality. Examination of the impact of taxation upon human capital accumulation may provide insight into the influence of taxes on effective labor supply.

Not much work has been done on integrating taxation into human capital analysis. Boskin (1975) studies a model of individual lifetime earnings maximization and shows that a proportional wage tax has no effect on human capital. James Heckman and John Driffill note that income (as opposed to earnings) taxes will change the optimal amount of human capital because they change net discount rates. These results are all obtained from models in which the returns to human capital investment are known with certainty. However, as David Levhari and Yoram Weiss emphasize, uncertainty is of paramount importance in human capital investment decisions. Although the taxation of nonhuman capital under uncertainty has been studied extensively (see James Tobin or Joseph Stiglitz), it will be explained below that results from this literature do not carry over to the human capital case. One of our main goals is to analyze theoretically the effect of taxation on human capital accumulation in a model which explicitly allows for uncertainty.

A closely related issue is the optimal taxation of earned income. Previous studies of efficient earnings taxation have treated returns to human capital as certain (see Nicholas Stern and Janusz Ordover). However, to the extent that introducing uncertainty changes behavioral responses to taxes, it may change optimal tax rates as well. A second major purpose of this paper, then, is to explore the problem of efficient taxation in the uncertainty model.

In Section I, we begin by studying a simple certainty model of human capital accumulation and contrast the tax implications with those of the corresponding uncertainty case. In the basic model with certainty and fixed labor supply, human capital investment is unaffected by a proportional earnings tax. This is because the tax reduces proportionately the costs of investment as

---

*Princeton University, and Princeton University and National Bureau of Economic Research, respectively. We would like to thank Alan Blinder, George Borts, Willem Buiter, Yoram Weiss, and a referee for useful suggestions. Part of Rosen's work on this project was completed while he was a visiting fellow at the Institute for Advanced Studies, Hebrew University of Jerusalem. Financial support was provided by NSF grant SOC-7907847. An extended version of this article is available as a memorandum of Princeton University's Industrial Relations Section with the same title.

[1] The analysis of the leisure-income choice in the presence of taxation is discussed carefully by Richard Musgrave (ch. 11) and much of the literature is surveyed by the Organization for Economic Cooperation and Development. Michael Boskin (1976) has critically reviewed empirical studies of the impact of taxes on hours of work.

[2] See the studies summarized by Glen Cain and Harold Watts (pp. 332–33). However, the labor supply response of other groups such as married women is considerably higher. See Cain and Watts (pp. 336–37) or Rosen.

well as the returns. Under uncertainty, however, the earnings tax affects human capital investment in two ways. First, it changes the riskiness of human capital. Second, it generates an income effect which may influence the individual's willingness to bear risk. After proving these propositions, we examine a number of extensions of the basic model, and ascertain how they influence the tax sensitivity of human capital accumulation.

Section II discusses the problem of optimal earnings taxation when the returns to human capital are uncertain. We show that under uncertainty the efficient earnings tax rate may exceed zero, even when lump sum taxation provides an alternative source of revenue. This finding contrasts, of course, with the well-known result from the optimal tax literature that, under certainty, lump sum taxation provides the most efficient means of raising revenue. The difference arises because the market fails to insure against the risk associated with human capital. The government, via the earnings tax, acts as insurer and thus increases welfare.

I. The Effect of Taxation on Investment in Human Capital

This section analyzes the effect of income and earnings taxation on the accumulation of human capital. Part A develops a model of human capital accumulation and savings under certainty and derives propositions on the effect of taxation on human capital investment. In Part B we consider the effects of taxation when the return on human capital is stochastic. To focus upon the effects of uncertainty, the decision to invest in human capital is treated in isolation from savings decisions and the labor-leisure choice. The effects of taxation on investment in human capital when savings and leisure are endogenous are considered respectively in Parts C and D.

A. *The Certainty Case*

Let us begin by analyzing human capital accumulation when its returns are certain. Although some of the certainty results have appeared before, it is useful to review them briefly, both to have a clear statement and to set the stage for the uncertainty case.

We assume that an individual's productive life is divided into two periods, each of which contains one unit of productive time. In the first period he allocates a proportion H of his productive time to the accumulation of human capital.[3] He receives no contemporaneous compensation for this effort either in the form of remuneration or utility. These foregone earnings are the only costs of accumulating human capital.[4] In the remaining $1-H$ share of his productive time in period 1 the individual earns a pretax wage rate w^1. Pre-tax labor income in the second period of his working life is given by $h(H)$, where $h' \geqslant 0$ and $h'' \leqslant 0$. Thus, human capital investment increases earnings capacity, but at a diminishing rate.[5]

In period 1, the worker earns an after-tax labor income of $(1-t)\ (1-H)w^1$, where t represents the tax rate on earned income. He consumes an amount C^1. Individuals are endowed at the beginning of their working lives with an amount W^1 of nonhuman wealth. Accretion to nonhuman wealth in the first period is thus $(1-t)\ (1-H)w^1-C^1$. In the second period, the worker earns after taxes $(1-t)h(H)$. In addition, nonhuman wealth at the beginning of period 2 is given by

$$(1) \quad W^2=(1+\phi r)\left[W^1+(1-H)\theta w^1-C^1\right]$$

Here $\theta \equiv 1-t$ and $\phi \equiv 1-\tau$, where τ denotes the rate at which interest income is taxed.

The worker's problem is to choose nonnegative C^1 and H to maximize the smooth, increasing and concave lifetime utility function $u(C^1, C^2)$, where

---

[3] Throughout the paper, H is assumed to be nonnegative—individuals cannot borrow human capital. We restrict our analysis to consider only interior solutions.

[4] This is a common assumption. See Heckman and especially Boskin (1975) for its justification. Although it would be possible also to allow for direct costs, this would complicate the analysis without adding any important insights.

[5] Contrast this to the problem of buying nonhuman capital, for which the gross rate of return is independent of the amount purchased. Levhari and Weiss emphasize this distinction and discuss its consequences.

$$C^2 \equiv \theta h(H) + W^2 \tag{2}$$

constitutes total resources available for consumption or bequests in period 2.

The first-order conditions for an interior maximum are

$$u_1 - (1+\phi r)u_2 = 0 \tag{3a}$$

$$\theta u_2\left[h' - (1+\phi r)w^1\right] = 0 \tag{3b}$$

Equation (3a) gives the usual condition that discounted marginal utility of consumption be equal in each period. Division of (3b) by $\theta U_2(>0)$ leads to the result that human capital investment occurs until the increase in the second-period wage from an additional investment of one hour is equal to the return on saved earnings from period 1. Optimal H is thus independent of both the utility function and the rate at which wages are taxed, a result very similar to that obtained by Boskin (1975).

Neutrality of H does not hold when period 2 labor supply is endogenous. Work effort in period 2 in effect constitutes the utilization of human capital. To the extent that earnings taxation influences labor supply, then, it changes the return on human capital. We illustrate this more precisely in Part D below.

Nor is H independent of τ. Differentiating (3b) with respect to τ one obtains

$$\frac{dH}{d\tau} = -\frac{rw^1}{h''} \tag{4}$$

a positive number; increased taxation of interest income encourages investment in human capital. Since a proportional tax on earnings has no impact on optimal H while a proportional interest tax increases H, a proportional income tax, which is just a combination of both, must increase H. This result is similar to one found by Heckman in the context of a life cycle model of human capital accumulation. It may at first seem counterintuitive that an income tax could actually increase H. However, as Heckman points out, this makes sense once it is realized that: a) an income tax depresses the after-tax rate of interest and, with deductability of interest payments, lowers the cost of borrowing; and b) foregone earnings costs of investment may be written off when incurred (p. 512). More intuitively, this is simply a substitution of a cheaper method of carrying consumption from period 1 to period 2.

We have assumed, as did Heckman, that the pre-tax interest rate is unaffected by taxation. However, Heckman formulates the human capital problem quite differently. He views human capital accumulation as augmenting the time endowment. In our model, the time endowment is constant and human capital increases the reward for spending time at work. Thus, in our model, investing in human capital in period 1 raises the opportunity cost of leisure in period 2, while in Heckman's it simply increases full income. Therefore, in our model, taxes on earnings which distort the human capital investment decision also distort the labor-leisure choice. In Heckman's model, earnings taxes affect the labor-leisure choice only via an income effect.

B. *A Simple Uncertainty Case*

Given the substantial uncertainty surrounding human capital decisions it is natural to ask how these results might change when the return to human capital investment is random.[6] The question of uncertainty in this context is of special interest since the risk associated with investment in human capital is typically neither insurable nor diversifiable. In many cases the market does not provide insurance against the vagaries of wage rates. Problems of moral hazard associated with insurance in general are especially pervasive in the insurance of human capital.[7] The private insurer, unable to distinguish clearly between external events and the endogenous behavior of the insured party, would provide an incentive for an insured worker to work less hard, spend less time seeking a higher-paying job, or otherwise earn an income below potential. In such a situation the market is unlikely to provide insurance.

---

[6] The impact of uncertainty upon human capital investment has been discussed by Levhari and Weiss. Their analysis ignores taxation.

[7] Kenneth Arrow discusses problems of moral hazard as they arise in insurance markets.

Let us assume that period 2 pre-tax wage income is given by $xh(H)$ where x is a random variable with a mean of unity and with support $[a, b]$, where $a \geqslant 0$. As Levhari and Weiss note, this is a somewhat restrictive way to model the randomness of human capital returns. Nevertheless, the multiplicative specification is a tractable one which yields interesting results.

We will also assume, for the moment, that period 1 consumption C^1 is given at $\bar{C}^1$. While this assumption does considerable violence to reality it simplifies the analysis and allows us to focus solely on the human capital decision. We study the interaction of the consumption and human capital accumulation decisions under uncertainty in Part C.

In period 1 the individual chooses an amount of human capital H to maximize

$$(5) \qquad \chi = E[u(C^1, C^2)]$$

where

$$(6) \quad C^2 = x\theta h(H) + (1+\phi r)[\theta w^1(1-H) + W^1 - \bar{C}^1]$$

The first-order condition for a maximum is given by

$$(7) \qquad E(u_2 F) = 0$$

where

$$(8) \qquad F \equiv x\theta h' - (1+\phi r)\theta w^1$$

is the difference between the expected return on human capital and the return on nonhuman capital.

Differentiating (7) with respect to t and solving for dH/dt, we find

$$(9) \quad \frac{dH}{dt} = E[u_{22}FI]/E[u_{22}F^2 + u_2\theta x h'']$$

where

$$(10) \quad I \equiv xh(H) + (1+\phi r)w^1(1-H)$$

is the sum of labor income in both periods expressed in period 2 units.

Diminishing marginal utility of C^2 and concavity of h, which guarantee satisfaction of the second-order condition for a maximum, imply the denominator of (9) is negative, while the sign of the numerator is in general ambiguous. The sign can, however, be inferred from information on preferences and initial wealth. The following lemma is used to make such inferences:

LEMMA: *Given a random variable x with p.d.f. $f(x)$ and support $[a, b]$ and given a function $g(x)$ with the property that there exists an $x^* \epsilon [a, b]$ such that $g(x) \geqslant 0$ for $x \geqslant x^*$ and $g(x) \leqslant 0$ for $x \leqslant x^*$, then, if*

$$\int_a^b g(x) f(x) dx = 0$$

the expression

$$(11) \quad \int_a^b \Gamma(x) g(x) f(x) dx > 0 \qquad (<0)$$

if $\Gamma(x)$ is uniformly increasing (decreasing). (A proof is provided in the Appendix.)

To apply the lemma we rewrite the numerator of expression (9) as

$$(12) \quad E[(u_{22}/u_2) I u_2 F] = -E[A I u_2 F]$$

where $A \equiv -u_{22}/u_2$, the degree of absolute risk aversion with respect to consumption in period 2. From (7), $E(u_2 F) = 0$. Furthermore, $u_2 F$ satisfies the criterion for $g(x)$ in the lemma that there exist an x^* such that for $x \leqslant x^*$, $u_2 F \leqslant 0$ while for $x \geqslant x^*$, $u_2 F \geqslant 0$. This is true since F is monotonically increasing in x while u_2 is always positive. At x^*, then, $F = 0$. Thus, via the lemma, expression (12) is positive or negative as $d(AI)/dx \lessgtr 0$. Since the denominator of expression (9) is negative, dH/dt is positive or negative as $d(AI)dx \gtrless 0$. We now demonstrate:

PROPOSITION 1: *If period 1 consumption is held constant, earnings taxation increases* *investment in human capital if absolute risk aversion is constant.*

Under constant absolute risk aversion (*CARA*) set $\Gamma = AI$ where A is now constant. Thus

$$\frac{d\Gamma}{dx} = A\frac{dI}{dx} = Ah(H) \geqslant 0 \tag{13}$$

and the result follows from the lemma.

PROPOSITION 2: *If period* 1 *consumption is held constant and if relative risk aversion is constant, earnings taxation increases or decreases investment in human capital as* $W^1 \gtreqless C^1$.

Under constant relative risk aversion (*CRRA*) set $\Gamma = RI/C^2$ where the degree of relative risk aversion with respect to period 2 consumption, $R \equiv -u_{22}C^2/u_2$, is constant. Then

$$\begin{aligned} \frac{d\Gamma}{dx} &= R\frac{d(I/C^2)}{dx} \\ &= Rh(H)(1+\phi r)(W^1 - C^1)/(C^2)^2 \end{aligned} \tag{14}$$

Since all the terms on the right-hand side of (14) are unambiguously positive except $(W^1 - C^1)$, $d\Gamma/dx$ has the same sign as $(W^1 - C^1)$, and Proposition 2 follows from the lemma.

Propositions 1 and 2 make sense intuitively. An increase in t reduces both periods' average wages in the same proportion, thus reducing C^2. At the same time, increasing t reduces the riskiness of period 2 labor income.[8] The second effect tends to increase investment in the risky asset (human capital), while the first, if absolute risk aversion is decreasing, acts to discourage investment in the riskier undertaking. If *CARA* obtains, the first effect is absent, because reductions in income do not reduce the individual's willingness to bear risk, and hence do not affect the human capital decision. Thus, only the second effect is operative, which accounts for the unambiguous sign of (13).[9] In the case of *CRRA*, if W^1 exceeds C^1, then an increase in t reduces C^2 by a smaller proportion than that by which it reduces period 2 wage income and the second effect dominates. The converse holds if W^1 is less than C^1.

Turning now to the effect of an increase in τ, the tax rate on interest income, we differentiate (7) with respect to τ to obtain

$$\frac{dH}{d\tau} = \frac{rE\left[u_2(-\theta w^1) + u_{22}FG\right]}{\chi_{HH}} \tag{15}$$

where

$$G \equiv \theta w^1(1-H) + W^1 - \bar{C}^1 \tag{16}$$

is nonhuman wealth at the *end* of period 1, and $\chi_{HH} \equiv \partial^2\chi/\partial H^2$. The first term of the numerator is negative, while the second is positive for decreasing absolute risk aversion and $G > 0$. The second-order condition for a maximum guarantees that the denominator is negative. As in the certainty case, interest taxation lowers the average return on nonhuman relative to human wealth and thereby encourages human capital accumulation. Thus, there is a direct correspondence between the positivity of $rE[u_2(-\theta w^1)]/\chi_{HH}$ and expression (4). By lowering overall after-tax income of net savers, however, the tax discourages investment in human capital, the riskier form of wealth. The interest tax, then, like the wage tax, generates a substitution effect which tends to encourage human capital formation and an income effect which discourages it.

In light of the ambiguity induced by the income effect, we consider the effect of an increase in interest taxation accompanied by a sufficient change in t to keep expected tax revenues constant. Expected total tax revenue in period 2 dollars is

$$T = tE(I) + \tau rG \tag{17}$$

---

[8] As t increases, the individual gains less from high realizations of x, but also loses less when low values occur. Thus, risk is decreased. The reasoning here is similar to that employed by Tobin in his analysis of the effect of proportional taxation of unearned income upon risk taking. Algebraically, if σ^2 is the variance of the pre-tax wage, then $\theta^2\sigma^2$ is the variance of the net wage, and the latter is less than the former since $\theta < 1$.

[9] Similar results have been found in the literature on portfolio composition. Musgrave suggests that under certain conditions corporate profit taxes may expand investment; also see Stiglitz.

Around the point $T=0$ the change in t which must accompany a change in τ to keep T constant is

$$(18) \qquad \left(\frac{dt}{d\tau}\right)\bigg|_{T=0} = -rG/E(I)$$

Now, the effect of a revenue-compensated increase in τ, near $t=\tau=0$, an optimal human capital accumulation is

$$(19) \qquad \frac{dH}{d\tau} + \frac{dH}{dt}\left(\frac{dt}{d\tau}\right)\bigg|_{T=0}$$

Substituting (9), (15), and (18) into (19) yields

$$(19') \quad \left(\frac{dH}{d\tau}\right)\bigg|_{T=0} = \frac{-rE(u_2 w^1) + rGE[u_{22}F(E(I)-I)]}{\chi_{HH}}$$

The first term of the numerator is negative while the second term is positive since $u_{22}F(E(I)-I) = -u_2 FA(E(I)-I)$ and, letting $\Gamma = A(E(I)-I)$ and applying the lemma, Γ is decreasing in x. Thus, replacing an earnings tax with an interest income tax yielding equal expected revenue has an ambiguous effect on investment in human capital under uncertainty. This result contrasts with the certainty result, that moving from an earnings to an income tax necessarily encourages investment in human capital. While, as in the certainty case, the interest tax reduces the expected return on nonhuman capital relative to human capital, the earnings tax, unlike the interest tax, makes human capital investments less risky, encouraging investment in human capital.

C. *Wage Taxation, Human Capital Accumulation, and Savings under Uncertainty*

We now relax the assumption that period 1 consumption is exogenous, giving us essentially the Levhari-Weiss model augmented with taxes. The variables H and C^1 are now chosen jointly to maximize $\chi = E[u(C^1, C^2)]$. First-order conditions for an interior maximum are

$$(20) \qquad \chi_C = E[u_1 - (1+\phi r)u_1] = 0$$

$$\chi_H = E[u_2 F] = 0$$

where $\chi_C \equiv \partial\chi/\partial C^1$ and, as before $\chi_H \equiv \partial\chi/\partial H$. Differentiating (20) totally with respect to t and appropriate substitution yields

$$(21) \qquad \frac{dH}{dt} = \left(\frac{dH}{dt}\right)\bigg|_{\bar{C}^1} \lambda + \chi_{CH}\chi_{Ct}/\Delta$$

where: (i) $dH/dt|_{\bar{C}^1}$ represents the effect of a change in t on H if C^1 is held constant (i.e., expression (9)); (ii) $\Delta \equiv \chi_{HH}\chi_{CC} - (\chi_{HC})^2$ which is positive if the second-order condition for a maximum is satisfied; and (iii) $\lambda \equiv \chi_{CC}\chi_{HH}/\Delta > 1$. Furthermore

$$(22) \qquad \chi_{Ct} = E[(1+\phi r)u_{22} I] < 0$$

$$(23) \qquad \chi_{CH} = (1+\phi r)E[Au_2 F] \leqslant 0$$

if absolute risk aversion is nonincreasing. Thus, if $(dH/dt)|_{\bar{C}^1} \geqslant 0$, then interaction with savings tends to increase the positive impact of earnings taxation upon human capital accumulation. If, however, $(dH/dt)|_{\bar{C}^1} < 0$, then permitting C^1 to be endogenous has an ambiguous effect on investment in human capital.

An intuitive explanation of this finding is based on the fact that if absolute risk aversion is diminishing, higher levels of C^2 increase the individual's tolerance of uncertainty in period 2 and therefore increase investment in H, *ceteris paribus*. In the model with fixed C^1, when the earnings tax is imposed the decrease in income must come *entirely* out of C^2. Allowing C^1 to be endogenous allows the individual to react to an earnings tax increase by raising C^2 at the expense of C^1 and increasing H. Hence, the second term of (21) is positive.

D. *Wage Taxation, Human Capital Accumulation, and Labor Supply under Uncertainty*

We now assume that the number of hours worked in period 2 is endogenous. Total time available for work in period 2 is one, and the proportion consumed as leisure is denoted L. The after-tax wage in period 2 is $\theta xh(H)$ so that period 2 labor income is $\theta xh(H)(1-L)$. The worker chooses L with knowledge of the realized value of x. (The worker must choose H and C_1 before he knows x, however.) His objective is to maximize the expected value of a utility function

$$(24) \qquad u(C^1, C^2, L)$$

which is concave and increasing in each argument.

Given values of C^1 and H, chosen prior to knowledge of x, and the value of x itself, the individual has only to choose a value of L. The first-order condition for a maximum is

$$(25) \qquad -u_2(C^1, C^2, L)x\theta h(H) + u_3(C^1, C^2, L)=0$$

Implicitly, (25) defines a leisure demand function $L(w^2, W^2)$, where $w^2 \equiv \theta xh(H)$ and W^2 is nonhuman wealth at the beginning of period 2 as defined in equation (1). Since $C^2 = \theta xh(H)(1-L)+W^2$, we may define the function

$$(26) \quad V(C^1, w^2, W^2) = \max_{L} u(C^1, C^2, L)$$

For simplicity revert to the assumption that $C^1 = \bar{C}^1$ and so remove C^1 as an argument of V. Then V is just the indirect utility function showing maximum utility *given* some choice of human capital investment.

As before, the individual allocates time in period 1 between investment in human capital and working for wage w^1. He selects a portion H of his time in period 1 for human capital investment to maximize expected utility $\psi \equiv E[V(w^2, W^2)]$. The first-order condition for a maximum is given by

$$(27) \quad \psi_H = E\left[V_1 xh' - V_2(1+\phi r)w^1\right] = 0$$

where $\psi_H \equiv \partial\psi/\partial H$, $V_1 \equiv \partial v/\partial w^2$, and $V_2 \equiv \partial V/\partial w^2$.

Substituting Roy's identity,[10] $V_1 = (1-L)V_2$, into (27) gives

$$(27') \qquad E(V_2 B) = 0$$

where

$$(28) \quad B \equiv xh'(H)(1-L) - (1+\phi r)w^1$$

The effect of an increase in t on optimal investment in human capital is given by totally differentiating (27′) with respect to H and t and rearranging to obtain

$$(29) \qquad dH/dt = -\psi_{Ht}/\psi_{HH}$$

where

$$\begin{aligned}(30\text{a}) \quad \psi_{Ht} = E\Big\{ & V_2 xh' \frac{\partial L}{\partial t}\Big|_{\bar{H}} + BV_2 L_W xh \\ & - V_{22} B\big[(1-L)xh \\ & + (1+\phi r)(1-H)w^1\big]\Big\}\end{aligned}$$

$$\begin{aligned}(30\text{b}) \quad \psi_{HH} = E\Big\{ & V_2\Big[(1-L)xh'' - \frac{\partial L}{\partial H} xh' \\ & - \theta xh' L_W B V_2 + V_{22} B^2\Big\}\end{aligned}$$

where

$$(31) \qquad L_W \equiv \partial L/\partial W^2$$

Under certainty $B=0$ and only the first term of (30a) is operative. For $\psi_{HH} < 0$, as implied by the second-order conditions, dH/dt has the opposite sign of $\partial L/\partial t|_{\bar{H}}$. Thus if earnings taxation reduces work effort it reduces investment in human capital. To determine whether or not this is likely to be the case, we note that

$$(32) \quad \frac{\partial L}{\partial t}\Big|_{\bar{H}} = -L_w xh - L_W(1-H)w^1$$

---

[10] Roy's identity relates the demand for a commodity to partial derivatives of the indirect utility function. Its interpretation when leisure is one of the commodities is discussed by Michael Abbott and Orley Ashenfelter.

A number of time-series studies[11] of labor supply suggest that L_w is close to zero. If such is the case, then the first term of (32) is zero, while the second term is negative if leisure is a normal good. Thus, hours of work increase with t and therefore there is a positive relationship between earnings taxation and investment in human capital in the certainty model with endogenous leisure.

Under uncertainty we must also examine the second and third terms of (30a). The third term of ψ_{Ht} is the analogue of the numerator of expression (9) and thus is ambiguous in sign. It is, as indicated by Propositions 1 and 2, positive under *CARA* or, if $W^1 > C^1$, under *CRRA*.[12] The second term of ψ_{Ht} is equal to

$$(33)\qquad L_W h(H)\int_a^b xV_2Bf(x)dx$$

which is positive if $L_W > 0$,[13] i.e., if leisure is normal.

Thus, introducing uncertainty tends to amplify the positive impact of an earnings tax on human capital investment under *CARA* or *CRRA* if $W^1 > C^1$. Conversely, if preferences are characterized by *CRRA* and the individual is a net borrower (i.e., if $W^1 < C^1$), then in the presence of uncertainty earnings taxes may discourage human capital formation even if they increase period 2 labor supply.

II. Optimal Taxation and Investment in Human Capital

In this section we consider the structure of an optimal linear earnings income tax with and without uncertainty in the return to human capital.[14] The two-period model developed in Section II serves as a starting point for the analysis. Throughout, wage uncertainty is viewed as a consequence of a risky return to human capital. However, it should be emphasized that the results would be equally valid if wage rates were fixed exogenously but stochastic.

The government's objective is to raise from each individual an expected per capita revenue $\bar{R}$ through a combination of earnings and lump sum taxes.[15] It wishes to maximize the expected after-tax utility of the individual. As in Section II, Part D, we assume that period 1 consumption is predetermined while hours worked in period 2 are subject to choice. We choose this formulation both for reasons of tractability and because much of the previous analysis of the excess burden of earnings taxation has focused on the labor-leisure question.[16] Finally, we assume that there is no pre-existing interest tax so that $\phi = 1$.

The government's problem, then, is to choose t and a per capita lump sum tax T to raise expected tax revenue

$$(34)\qquad \bar{R} = t\left[(1-L)\bar{x}h(H) + (1+r)(1-H)w^1\right] + T$$

in such a way as to maximize

$$(35)\qquad \psi = E\left\{V\left[\theta xh(H), (1+r)\cdot\left((1-H)\theta w^1 + W^1 - \bar{C}^1\right) - T\right]\right\}$$

where $V(\cdot)$ is the indirect utility function defined in Section II, Part D. The first-order condition for a maximum is [17]

$$(36)\qquad \frac{d\psi}{dt} = E\left\{V_1\frac{dw^2}{dt} + V_2\left(\frac{dW^2}{dt} - \frac{dT}{dt}\right) + \right.$$

---

[11] See, for example, Robert Lucas and Leonard Rapping. In addition, cross-section studies of prime age males also suggest negligible wage elasticities. See the essays in Cain and Watts.

[12] Here we define *CARA* and *CRRA* in terms of derivatives of the indirect utility function with respect to W^2.

[13] The expression inside the integral is positive as may be shown by applying the lemma, letting $\Gamma(x) = x$.

[14] Peter Diamond, Jay Helms, and James Mirrlees study optimal redistributive taxation in an uncertainty context, but do not consider explicitly the human capital aspects of the problem.

[15] Our assumption that the government is only concerned with expected revenue need not imply that the government is risk neutral. If the risk associated with human capital investment is independent across individuals and if the number of individuals is large, then the law of large numbers will guarantee the government a constant total revenue despite uncertainty on the individual level. The government is, in this case, simply a more efficient risk pooler than the individual.

[16] See, for instance, Mirrlees and Ray Fair.

[17] Here $(dH/dt)|_{\bar{y}}$ and $(dL/dt)|_{\bar{y}}$ represent the *expected*-income-compensated effects of an increase in t on H and L, respectively.

$$\left[V_1\frac{\partial w^2}{\partial H}+V_2\left(\frac{\partial W^2}{\partial H}-\frac{\partial T}{\partial H}\right)\right]$$

$$\left(\frac{dH}{dt}\right)\Big|_{\bar{y}}-V_2\frac{\partial T}{\partial L}\left(\frac{dL}{dt}\right)\Big|_{\bar{y}}\Bigg\}=0$$

The individual's problem is to choose H to maximize (35) taking T and t as given. The first-order condition for this maximum is that

$$E\left[V_1xh'-V_2(1+r)w^1\right]=0 \tag{37}$$

Substituting (34), (35), (37), and Roy's identity into (36) gives

$$\begin{aligned}&-E\left[V_2(1-L)h(H)(\bar{x}-x)\right]\\&+tE\left\{\left[V_2\left[(1-L)\bar{x}h'(H)-(1+r)w^1\right]\right\}\right.\\&(dH/dt)|_{\bar{y}}+tE\left[V_2\bar{x}h(H)\right](dL/dt)|_{\bar{y}}=0\end{aligned} \tag{38}$$

Subtracting $0=E\{V_2[(1-L)xh'(H)-(1+r)w^1]\}$ from (38) allows us to rewrite (38) as

(39)

$$\begin{aligned}\frac{d\psi}{dt}&=\left[h(H)+th'(H)(dH/dt)|_{\bar{y}}\right](1-L)E\\&\left[V_2(\bar{x}-x)\right]-tE\left[V_2\bar{x}h(H)\right](dL/dt)|_{\bar{y}}=0\end{aligned}$$

Under certainty $x=\bar{x}$, and (39) is satisfied at $t=0$. Not surprisingly, lump sum taxation is efficient under certainty. However, when the return to human capital is uncertain, ψ is not maximized at the point $t=0$. Indeed, we will show that under reasonable conditions $d\psi/dt$ is positive at this point, implying that at least partial reliance on a wage tax is more efficient than a pure lump sum tax.

Under uncertainty, near the point $t=0$, expression (39) has the same sign as $E[V_2(\bar{x}-x)]$. In order to sign this expression, we note that

$$E\left[\bar{x}-x\right]=0 \tag{40}$$

and apply the lemma. To do so requires that we determine whether $dV_2/dx(=V_{12}\theta h(H))\gtreqless 0$. But differentiating Roy's identity with respect to W^2 and rearranging yields

$$V_{12}=-L_WV_2+(1-L)V_{22} \tag{41}$$

The first term of (41) is negative as long as the demand for leisure is normal while risk aversion implies that the second term is negative as well. An increase in x, then, tends to lower the marginal utility of nonhuman wealth.

Therefore, by the lemma, at $t=0$ expression (39) is positive. Welfare is increased by *raising* t and lowering T to maintain expected revenue constant. In contrast with the certainty case, sole reliance upon lump sum taxation is inefficient. Intuitively, allocative efficiency is enhanced by a tax which reduces the riskiness of investment in human capital. Through the wage tax the government insures the risk-averse individual's uncertain undertaking, and thus compensates partially for the nonexistence of a market in human capital insurance. Furthermore, note that this result is independent of the effect of earnings taxation on investment in human capital or on hours or work.

Since near $t=0$ taxing wages and returning the revenue as a lump sum enhances expected welfare,[18] it might be argued that full government insurance of human capital investment via a 100 percent wage tax is optimal.[19] However, such a result is unlikely. Since $(dL/dt)|_{\bar{y}}$ represents the expected-income-compensated effect on leisure of an increase in the wage tax, it must be positive in sign. The negativity of the last term of (39) near $t=1$ is likely to make the sign of (39) negative in this region. Thus, efficient taxation requires wage taxation at a level strictly between 0 and 100 percent accompanied by lump sum taxes or transfers to meet revenue requirements.

---

[18] A similar exercise can be performed to determine the optimal taxation of interest income. Since in this model the return to nonhuman capital is certain, the interest tax only falls on nonstochastic components of income and does not reduce the riskiness of human capital. It can be shown that when lump sum taxation is feasible, the optimal interest tax is zero whether or not the return to nonhuman capital is uncertain.

[19] A result somewhat similar to this has been obtained by Joram Mayshar, who studies the subsidization of risky private projects by the government.

This result suggests the value of additional theoretical analysis of the design of efficient tax systems. The familiar theorem that, under certainty, commodity or income taxation is necessarily second best does not generalize to situations characterized by the presence of uncertainty which is neither diversifiable nor insurable. The uninsurability of most investments in human capital suggests that departures from lump sum taxation may be optimal more frequently than previously realized.

Appendix

PROOF of the lemma that appears in Section II, Part B. This proof is very similar to one provided by Levhari and Weiss (p. 958).

Note that

$$\int_a^{x^*}\Gamma(x)g(x)f(x)dx$$

$$\gtrless\Gamma(x^*)\int_a^{x^*}g(x)f(x)dx\leqslant 0$$

and $\int_{x^*}^{b^*}\Gamma(x)g(x)f(x)dx$

$$\gtrless\Gamma(x^*)\int_{x^*}^{b}g(x)f(x)dx\geqslant 0$$

as $\Gamma(x)$ is monotonically increasing or decreasing. Thus

$$\int_a^{b}\Gamma(x)g(x)f(x)dx$$

$$\gtrless\Gamma(x^*)\int_a^{b}g(x)f(x)dx=0$$

as $\Gamma(x)$ increases or decreases monotonically.

REFERENCES

Kenneth J. Arrow, *Essays in the Theory of Risk Bearing*, Amsterdam 1970.

M. Abbott and O. Ashenfelter, "Labour Supply, Commodity Demand, and the Allocation of Time," *Rev. Econ. Stud.*, Oct. 1976, *43*, 389–411.

M. J. Boskin, "Notes on the Tax Treatment of Human Capital," Nat. Bur. Econ. Res. work. paper no. 116, Stanford 1975.

——, "On Some Recent Econometric Research in Public Finance," *Amer. Econ. Rev. Proc.*, May 1976, *66*, 102–09.

Glen Cain and Harold Watts, *Income Maintenance and Labor Supply*, Chicago 1973.

P. Diamond, J. Helms, and J. Mirrlees, "Optimal Taxation in a Stochastic Economy: A Cobb-Douglas Example," mimeo., M.I.T., 1978.

J. Driffill, "A Simulation Analysis of the Effects of Taxes," mimeo., Univ. Southampton 1977.

R. C. Fair, "The Optimal Distribution of Income," *Quart. J. Econ.*, Nov. 1971, *85*, 551–79.

J. Heckman, "A Life-Cycle Model of Earnings, Learning, and Consumption," *J. Polit. Econ.*, Aug. 1976, *84*, Part II, S11–S44.

D. Levhari and Y. Weiss, "The Effect of Risk on the Investment in Human Capital," *Amer. Econ. Rev.*, Dec. 1974, *64*, 950–63.

R. Lucas and L. Rapping, "Real Wages, Employment and Inflation," in Edmund Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

J. Mayshar "Should Government Subsidize Risky Private Projects?," *Amer. Econ. Rev.*, Mar. 1977, *67*, 20–28.

J. G. Mirrlees, "An Exploration in the Theory of Optimum Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, 179–208.

Richard A. Musgrave, *The Theory of Public Finance*, New York 1959.

J. A. Ordover, "Distributive Justice and Optimal Taxation of Wages and Interest in a Growing Economy," *J. Publ. Econ.*, Jan./Feb. 1976, *5*, 139–60.

H. Rosen, "Taxes in a Labor Supply Model with Joint Wage-Hours Determination," *Econometrica*, June 1976, *44*, 485–507.

N. H. Stern, "On the Specification of Models of Optimum Income Taxation," *J. Publ. Econ.*, July/Aug. 1976, *5*, 123–62.

J. Stiglitz, "The Effects of Income, Wealth and Capital Gains Taxation on Risk Taking," *Quart. J. Econ.*, May 1969, *83*, 263–283.

J. Tobin, "Liquidity Preference as Attitude Toward Risk," *Rev. Econ. Stud.*, Feb. 1958, *25*, 65–86.

Organization for Economic Cooperation and Development, *Theoretical and Empirical Aspects of the Effects of Taxation on the Supply of Labour*, Paris 1975.

Efficient Protection under Uncertainty

By ELHANAN HELPMAN AND ASSAF RAZIN*

Suppose that in a small trading country a policymaker desires to increase the allocation of productive resources to the import-competing sector beyond the free trade allocation. It is well known that in a deterministic environment the first best policy is to subsidize the output of the import-competing sector, while a tariff (specific or ad valorem), an export tax (specific or ad valorem), an import quota, and an export quota are all equivalent second best policy measures. How do these policy instruments compare in the presence of uncertainty? To the best of our knowledge, this question has not been addressed in the literature.[1]

For the purpose of this study, we employ the general equilibrium framework with financial markets that has been developed in our book. Using *expected utility* as the welfare criterion, this framework enables us to compare output subsidies, trade taxes, and quotas as alternative means of protection. In addition, it enables us to consider interventions in financial capital markets—like equity taxes and subsidies—as means of protection. Finally, it allows us to consider the relevance of the existence of international trade in securities to these policy rankings.

In the type of economy that we consider, the allocation of productive resources is not governed directly by commodity prices (as in a deterministic environment) but rather by equity prices—with equity prices *depending*, of course, on commodity prices. Therefore we will work out precisely how policies directed towards commodity markets affect equity prices. For tariffs, this has been done in our 1978a paper. Here we also develop this link for quotas. This issue is of some independent interest, because it is the first demonstration (to the best of our knowledge) showing how import licenses are valued on the stock market.

With regard to intervention in commodity markets, the presence of uncertainty enlarges the policy menu. Tariff rates as well as quotas can be state dependent or state independent. We feel that state-independent policies are more relevant, because from a practical point of view it is hard to see how a government can implement a state-dependent protective policy, such as applying a tariff rate that depends on inches of rainfall. For this reason, most of our analysis concentrates on state-independent policies.

We show that in the absence of international trade in securities a subsidy to sales of equities of the import-competing sector is the first best policy, and that this policy is equivalent to a state-independent rate of subsidy on output sales of this sector. Other state-independent policies cannot in general be ranked. We show that a state-independent ad valorem tariff may dominate a state-independent import quota, and vice versa, depending on the risk structure and attitudes towards risk.

In the presence of international trade in securities, we show that in the absence of restrictions on short sales of equities, a subsidy to sales of equities of the import-competing sector dominates a state-independent ad valorem tariff. In the presence

---

*Tel-Aviv University. We are grateful to the Ford Foundation for financial support and to the University of Rochester for its hospitality during our work on this paper. Finally we would like to thank George H. Borts and a referee for their comments.

[1] There are studies which compare tariffs and import quotas under uncertainty as alternative means to achieve certain objectives, see, for example, Gideon Fishelson and Frank Flatters, and Partha Dasgupta and Joseph Stiglitz, but their objectives do not include protection of the import-competing industry. The former study is concerned with the exploitation of monopoly power in foreign trade while the latter is concerned with raising a given level of *expected* revenue (a quota raises revenue since the authors assume that import licenses are auctioned off *ex post*). In both cases, *expected* consumer's surplus is used to measure the economy's welfare level.

of short-sales restrictions, a tariff may dominate an equity subsidy if investors choose to specialize in equity holdings of the exporting sector.

We present the framework of analysis in Section I. In Section II we describe equilibria under alternative policy measures. First best policies are discussed in Section III. In Section IV we compare tariffs with quotas in the absence of international trade in securities. Finally in Section V we review our main conclusions, providing a detailed discussion of other possible means of protection.

I. The Framework

We employ the two-sector static model that was developed in our book (see also our 1978a paper). In this model sector j produces its output according to

$$(1)\qquad Q_j(\alpha)=\theta_j(\alpha)f_j(L_j,K_j)\qquad \alpha=1,2,\ldots,S;\, j=1,2$$

where L_j = labor employed by sector j, K_j = capital employed by sector j, f_j is a linear homogeneous neoclassical production function, $\theta_j(\alpha)$ is a random variable which may obtain different values in different states of nature α, $\alpha=1,2,\ldots,S$, and $Q_j(\alpha)$ is output of sector j in state α.

The choice of factor inputs is made by firms before the realization of a state α, and this is done so as to maximize the firms' net value on the stock market. The firm's net value equals the value of its initial equities. We assume that factor costs are financed either directly by the *initial* stockholders or by issuance of new equities. In this case the aggregate return to final stockholders of sector j is the value of output $p_j(\alpha)\theta_j(\alpha)f_j(L_j,K_j)$, where $p_j(\alpha)$ is the price of good j in state α. We call $f_j(\cdot)$ the number of *real equities* produced by sector j. Using this definition, the final holder of share s_j of sector j's equities holds in effect $z_j=s_jf_j(\cdot)$ units of its real equities. A real equity entitles its holder to the return $p_j(\alpha)\theta_j(\alpha)$ in state α, which is the value of $\theta_j(\alpha)$ units of good j in state α. It is clear that there is a one-to-one correspondence between holdings of shares and holdings of real equities. Given aggregate labor and capital endowments, the hypotheses of net value maximization and competitive factor-market clearing enable one to derive sectoral real equity supplies $Z_j(q_p)$, $j=1,2$, which depend on the relative producer price of the second real equity q_p, with $Z_1'(q_p)<0$ and $Z_2'(q_p)>0$ (except for corners in which equalities hold).

Using the supply functions $Z_j(q_p)$ one can trace out (by varying q_p) a transformation curve between the outputs of real equities, like TT in Figure 1. At each point on this curve the marginal rate of transformation is just equal to q_p so that

$$(2)\qquad Z_1'(q_p)+q_pZ_2'(q_p)=0$$

where a prime denotes a derivative.

The representative consumer-investor of our economy engages in a two-stage decision process. First, before the resolution of uncertainty, he trades in the stock market ending up with holdings of z_j units of real equities of type j, $j=1,2$. His net trading in the stock market is financed by the proceeds from sales of labor and capital services plus any transfers which may occur from the government. Second, after the resolution of uncertainty, he uses the aggregate return on his portfolio plus any additional income (like government transfers) to purchase consumption goods. In the second stage, given the realization of state α, the individual solves the problem

$$(3)\qquad \text{Choose } c_1(\alpha),c_2(\alpha)\geqslant 0$$

$$\text{to maximize}\quad u[c_1(\alpha),c_2(\alpha)]$$

$$\text{subject to } c_1(\alpha)+\bar{p}(\alpha)c_2(\alpha)\leqslant\theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2+T(\alpha)$$

where $c_j(\alpha)$ = consumption of good j in state α, $u(\cdot)$ is a von Neumann-Morgenstern concave (i.e., risk-averse) utility function, $\bar{p}(\alpha)$ = the *domestic* relative price of good 2,

$[\bar{p}(\alpha)=\bar{p}_2(\alpha)/\bar{p}_1(\alpha)]$, and $T(\alpha)$ is the component of income in state α which is not due to real equity holdings; this component will be specified later. Observe that the first two terms on the right-hand side of the budget constraint represent the return on real equity holdings.

The solution to (3) yields the indirect utility function

$$v[\bar{p}(\alpha);\theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2+T(\alpha)]$$

which is concave in its second argument due to the assumption of risk aversion on the part of the consumer-investor. Now we can describe the individual's first-stage choice problem as the portfolio choice subject to an assets budget constraint, so as to maximize the expected value of the indirect utility function

$$\text{(4)} \qquad \text{Choose } z_1, z_2 \geqslant 0$$

to maximize

$$Ev[\bar{p}(\alpha);\theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2+T(\alpha)]$$

$$\text{subject to } z_1+q_c z_2 \leqslant Z_1(q_p)+q_p Z_2(q_p)+T$$

where $E=$ the expectations operator, based on subjective probability beliefs, with *correct* state-dependent expectations concerning $\bar{p}(\alpha)$, $\theta_1(\alpha)$, $\theta_2(\alpha)$, and $T(\alpha)$; $q_c=$ the investor's relative price of real equity 2; and T equals government transfers before the realization of state α. Due to the concavity of $v(\cdot)$ in its second argument, the assets indifference curves are convex to the origin, as drawn in Figure 1. In some limiting cases they may be straight lines.

The equilibrium relationship between commodity and equity prices can be seen from the first-order conditions of problem (4), which imply

$$q_c=Ev_I[\bar{p}(\alpha);\theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2$$
$$+T(\alpha)]\bar{p}(\alpha)\theta_2(\alpha)\div Ev_I[\bar{p}(\alpha);\theta_1(\alpha)z_1$$
$$+\bar{p}(\alpha)\theta_2(\alpha)z_2+T(\alpha)]\theta_1(\alpha)$$

where $v_I[\cdot]$ is the marginal utility of spending. This condition is in effect the condition of tangency between an assets indifference curve and an assets budget line. It is seen from this that commodity prices are related to real equity prices by means of two channels: 1) the marginal utility of spending $v_I[\cdot]$; and 2) the return on the second real equity $\bar{p}(\alpha)\theta_2(\alpha)$. Note that in the case of certainty, with $\theta_1(\alpha)=\theta_2(\alpha)=1$, this condition implies $q_c=\bar{p}$.

We assume that our economy is small, so that the foreign relative price of commodity 2, $p(\alpha)$, depends only on the state α. Domestic relative commodity prices $\bar{p}(\alpha)$, $\alpha=1,2,\ldots,S$, depend on domestic policies like tariffs or quotas. The above formulation is general enough to accommodate each of these policies.

Relative real equity prices depend on government policies in the capital market, and on the existence of international trade in securities. If there is international trade in securities, real equity prices are also assumed to be given to the economy (the small-country assumption in security markets), while in the absence of trade in securities, real equity prices are domestically determined so as to clear the stock market. Government intervention in security markets may create a wedge between relative real equity prices faced by domestic producers and investors, as well as—in the presence of trade in real equities—a wedge between foreign and domestic relative prices of identical real equities. Since we concentrate in this paper on three types of policy—tariffs, quotas, and security taxation—it is useful to describe at this stage the equilibrium of the economy in the presence of each type of these policies.

II. Equilibrium under Alternative Policies

The policies that we consider are directed toward protection of the import-competing industry; the purpose is to induce a resource flow from the export to the import-competing industry. For this purpose we assume that the import-competing industry is well defined in the sense that the same good, say good 2, is imported in every state.

A. *Equity Subsidy*

An equity subsidy is a subsidy to the *domestic* issues of equities. The subsidy is given to type-2 real equities; that is, equities of the import-competing sector. In this case the producer's relative price of real equity 2 is equal to the investor's relative real equity price plus the subsidy; i.e., $q_p=(1+s)q_c$, $s>0$. The subsidy is financed by means of a lump sum tax (which the investor does not relate to his purchase of securities) before the realization of a state of nature. The lump sum tax is $-T$, where

$$T=-sq_cZ_2(q_p) \tag{5}$$

Since there is no intervention in commodity markets, domestic commodity prices equal foreign commodity prices, and state-dependent transfers equal zero.

$$\bar{p}(\alpha)=p(\alpha) \qquad \alpha=1,2,\ldots,S \tag{6}$$

$$T(\alpha)=0 \qquad \alpha=1,2,\ldots,S \tag{7}$$

Now we can describe the equilibrium in the stock market. In the presence of international trade in securities, we assume that foreigners are willing to exchange real equities at a constant relative price q $(=q_2/q_1)$. We also assume in this case, and this is done only for the sake of simplicity, that the domestic equities are the only available *types* of securities in the world. Thus, foreign securities have the same distributions of returns as domestic type-1 and domestic type-2 real equities.[2] In this case domestic investors face the foreign relative security price $q_c=q$, while firms face the relative price $q_p=(1+s)q$. The stock market equilibrium is described in Figure 1. Net value-maximizing firms choose point E_{sp} at which a line with the slope $(1+s)q$ is tangent to the real equities transformation curve TT. Without the subsidy they would have chosen a point on TT to the left of E_{sp}, since they would have equated q to the slope of TT. Hence, the subsidy induces a shift of resources into sector 2.

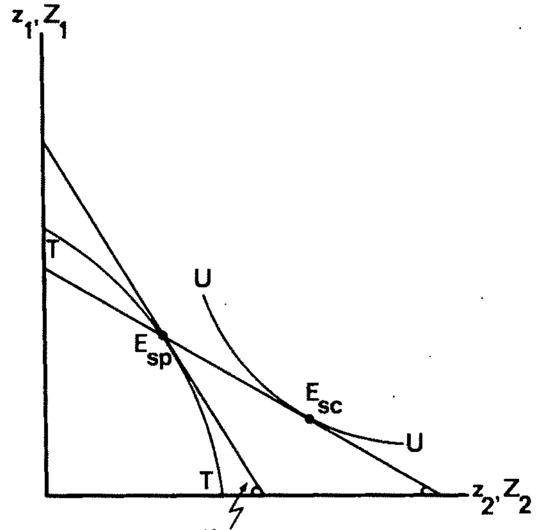


FIGURE 1

From the assets budget constraint in (4), using (5) and $q_c=q$, one can see that the investor's assets budget line has slope q and passes through E_{sp}. The investor chooses a portfolio at the point of tangency between the assets budget line and one of his assets indifference curves. This happens at point E_{sc}, with UU representing the highest achievable assets indifference curve.[3] An asset indifference curve is defined in the present case by all combinations of (z_1, z_2) which satisfy

$$Ev[p(\alpha);\theta_1(\alpha)z_1+p(\alpha)\theta_2(\alpha)z_2] = \text{constant} \tag{8}$$

If there is no international trade in securities, then given the subsidy rate s, the

---

[2] Our proofs also go through in the case in which foreigners issue securities with distributions of returns that differ from domestic real equities, as long as the assumption of smallness of the domestic economy is maintained. What is required is that the home country be a price taker in security markets, including prices of domestic real equities. However, in this more general case the simple diagrammatical exposition has to be abandoned. These remarks apply, of course, only to the case in which there is international trade in securities.

[3] In Figure 1, type-2 real equities are imported. It is, however, possible to have an equilibrium in which type-1 real equities are imported. The subsidy may also reverse the pattern of real equity trade.

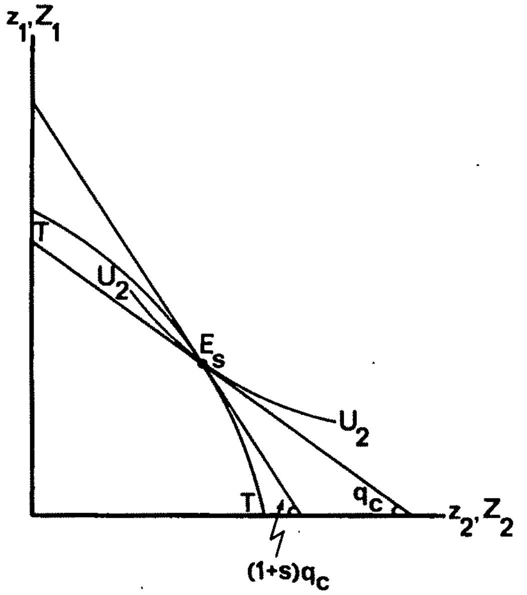


FIGURE 2

domestic relative price of real equity 2 faced by investors is domestically determined so as to clear the domestic stock market. In this case, the production point E_{sp} and the portfolio point E_{sc} coincide. Such an equilibrium is presented by point E_s in Figure 2. The set of assets indifference curves in Figure 2 is the same as in Figure 1. Observe also that without the subsidy more resources would be allocated to sector 1.

B. *Tariff*

Consider an ad valorem tariff with the same rate applying to all states of nature; the tariff rate is chosen before the realization of a state of nature. We believe this to be the relevant case, since it is difficult to conceive of a tariff policy in which the tariff rate is adjusted *ex post* to suit the realized state of nature. A specific tariff would be an appropriate alternative, but it causes difficulties in the presence of international trade in securities. By changing the level of the specific tariff, one changes the structure of returns on equities issued by the import-competing industry, and this affects the price of its equities in a way which can be worked out only in an explicit two-country model, a complication which we prefer to avoid.

In the presence of an ad valorem tariff at rate t, domestic relative commodity prices are

$$(9) \qquad \bar{p}(\alpha)=(1+t)p(\alpha) \quad \alpha=1,2,\ldots,S$$

We assume that tariff proceeds are redistributed back to consumers. Since tariff revenues occur after a realization of a state of nature, the redistribution of tariff proceeds also occurs after the realization of a state of nature. The required lump sum transfers are implicitly defined by

$$(10) \quad T(\alpha)=tp(\alpha)\{c_2[(1+t)p(\alpha);\theta_1(\alpha)z_1 + (1+t)p(\alpha)\theta_2(\alpha)z_2+T(\alpha)]-\theta_2(\alpha)Z_2\} \qquad \alpha=1,2,\ldots,S$$

In the presence of tariffs, the map of assets indifference curves depends on the tariff rate (see our 1978a paper), and it differs from the map of assets indifference curves defined by (8). In the present context the investor's assets indifference curve is defined by all combinations of (z_1, z_2) which satisfy

$$(11) \quad Ev[(1+t)p(\alpha);\theta_1(\alpha)z_1 + (1+t)p(\alpha)\theta_2(\alpha)z_2+T(\alpha)]=\text{constant}$$

where $T(\alpha)$ is kept at a predetermined level for every α. In order to represent an equilibrium using these indifference curves, $T(\alpha)$ must be kept at its equilibrium level.

Now, in the presence of a tariff and no intervention in financial markets, the investor's and firm's relative real equity prices coincide (i.e., $q_c=q_p$) and *ex ante* transfers equal zero (i.e., $T=0$). In the presence of international trade in real equities, the tariff raises the domestic relative price of real equity 2 by a rate which equals the rate of tariff (see our 1978a paper). The resulting stock market equilibrium is described in Figure 3. Net value-maximizing firms choose the resource allocation described by point E_{sp} at which the slope of TT is $(1+t)q$.

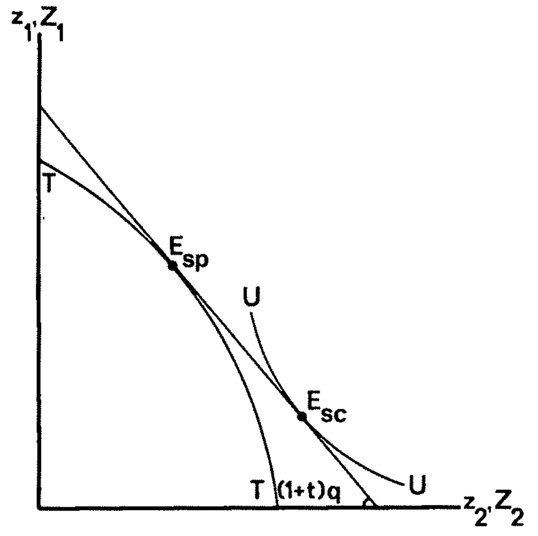


FIGURE 3

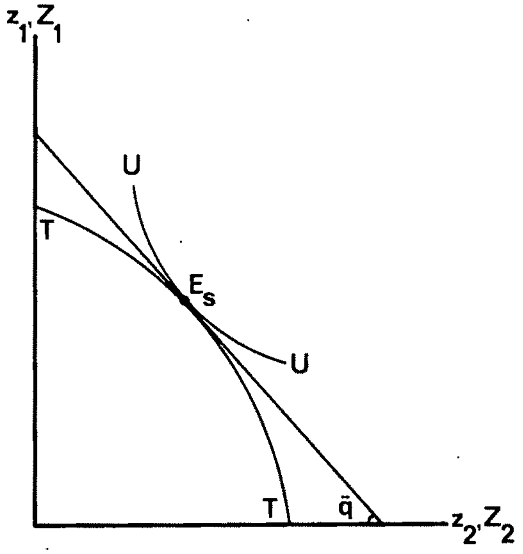


FIGURE 4

Investors choose a portfolio described by point E_{sc}. Without the tariff, firms would have chosen a point on TT to the left of E_{sp} at which the slope of TT is q. Hence, the tariff induces a resource flow into the second sector.

In the absence of international trade in securities, the domestic relative price of real equity 2 is determined so as to clear the stock market. The resulting equilibrium is described in Figure 4. The equilibrium relative price of real equity 2 is $\bar{q}$, which equals the slope of TT and UU at their common point of tangency, E_s. The assets indifference curve UU in Figure 4 does not belong to the same family of indifference curves as UU in Figure 3, because the state dependent transfers $T(\alpha)$ in Figure 4 differ from those in Figure 3.

We have shown in our 1978b paper that in this case a small tariff induces a resource flow into the import-competing sector if both goods are normal in consumption, but for large tariff rates, it is not clear which way resources will flow. We assume for the present purpose that ad valorem tariffs are protective.[4]

C. *Import Quota*

Our discussion of an import quota is restricted to the case in which there is no international trade in securities. The reason is similar to that which prevented us from discussing a specific tariff—the imposition of an import quota changes the pattern of returns on real equities that are issued by the import-competing sector. We cannot deal satisfactorily in a single country model with the effect on the price of this security even when the country under consideration is small in the world's financial capital markets.

We assume that the government imposes an import quota M, and the decision on the size of the quota is made before the realization of a state of nature. Thus, the same quota applies to all states and is binding in at least some of them. If the quota is sufficiently small, it will be binding in all states, but we do not require it to be of this type. The size of the quota affects domestic relative commodity prices in states in which the quota is binding; in states in which the quota is not binding, domestic relative commodity prices are $p(\alpha)$—world relative prices. Figure 5 describes an equilibrium in commodity markets in a particular state α.

---

[4] If a tariff is not protective, we can use a subsidy on imports in order to induce the desired resource flow (i.e., $t<0$). Our conclusions do not depend on this possibility.

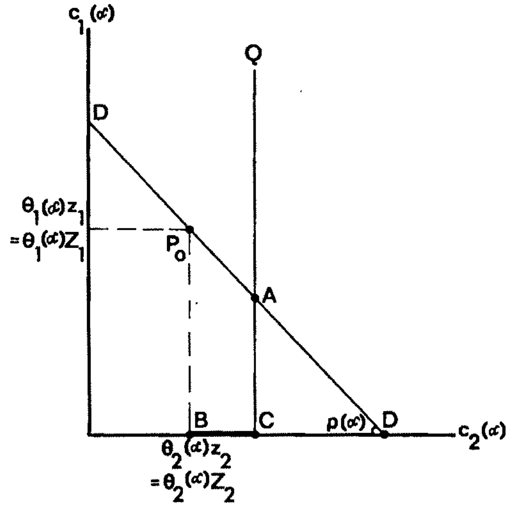


FIGURE 5

Point P_o describes a feasible consumption point which consists of the domestic output of goods. The straight line DD with slope $p(\alpha)$ that passes through P_o describes feasible consumption points which are achievable through foreign trade. The vertical line Q describes the quota constraint whose size is BC. Now, if an indifference curve is tangent to DD to the left of A, the quota is not binding, and the domestic relative price $\bar{p}(\alpha)$ equals $p(\alpha)$. If, however, an indifference curve is tangent to DD to the right of A, the quota is binding, and the domestic relative price $\bar{p}(\alpha)$ equals the slope of the indifference curve at point A, which is larger than $p(\alpha)$.

It is clear from here that domestic relative commodity prices depend, at least in some states, on the size of the quota. Since asset indifference curves depend, among other things, on the distribution of domestic commodity prices, the quota affects the structure of the asset indifference field. In the present setup an asset indifference curve is defined by all combinations of (z_1, z_2) which satisfy

$$(12)\quad E[\bar{p}(\alpha);\theta_1(\alpha)z_1 + \bar{p}(\alpha)\theta_2(\alpha)z_2 + T(\alpha)] = \text{constant}$$

where $T(\alpha)$ is implicitly defined by

$$(13)\quad T(\alpha) = [\bar{p}(\alpha) - p(\alpha)] \times \{c_2[\bar{p}(\alpha);\theta_1(\alpha)z_1 + \bar{p}(\alpha)\theta_2(\alpha)z_2 + T(\alpha)] - \theta_2(\alpha)Z_2\} \qquad \alpha = 1,2,\dots,S$$

Observe that (13) implies that $T(\alpha)=0$ in a state α in which the quota is not binding, and $T(\alpha)=[\bar{p}(\alpha)-p(\alpha)]M$ in a state α in which the quota is binding.

In our formulation, $T(\alpha)$ equals the value of the import licenses in state α. This can be interpreted as follows. Suppose that after a realization of a state of nature, the government auctions off the quota and redistributes the proceeds to consumers as lump sum transfers. These transfers equal $T(\alpha)$ in state α. Importing firms that buy import licenses have zero profits in every state α and their value on the stock market is zero. In this case, the representative investor's asset budget constraint is that given in (4), with $q_c = q_p$ and $T=0$, and stock market equilibrium is as described in Figure 4 with the indifference field defined by (12). Remember, however, that the indifference field under the quota differs from the indifference field under the tariff.

There is another interpretation which can be given to this equilibrium. Suppose that the government distributes the quotas to importing firms which are locally owned, and that the shares of these firms are traded on the domestic stock market. A license to import one unit of good 2 provides a return of $\bar{p}(\alpha)-p(\alpha)$ in state α, $\alpha=1,2,\dots,S$. If the quota is binding in state α, $\bar{p}(\alpha)-p(\alpha)>0$, and it pays to take full advantage of the import volume permitted by the license. If the quota is not binding in state α, $\bar{p}(\alpha)-p(\alpha)=0$, so that the firm's profits are zero even if its import volume should reach the limit permitted by the license. Hence, the vector of returns on a unit import license is

$$[\bar{p}(1)-p(1), \bar{p}(2)-p(2), \dots, \bar{p}(S)-p(S)]$$

and we have multiplicative uncertainty in the importing firms.

We may rewrite now the investor's problem (4) as

$$(4') \qquad \text{Choose } z_1, z_2, z_m \geqslant 0$$

to maximize

$$Ev\{\bar{p}(\alpha);\, \theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2 + [\bar{p}(\alpha)-p(\alpha)]z_m\}$$

subject to

$$z_1+\bar{q}z_2+q_m z_m \leqslant Z_1(\bar{q})+\bar{q}Z_2(\bar{q})+q_m M$$

where z_m = real equities of the importing firms held by the investor, q_m = the relative price of a real equity of an importing firm, and M = the number of real equities of importing firms available in the economy, which equals the import quota.

In equilibrium, $\bar{q}$ and q_m settle on market-clearing levels such that the solution to (4′) yields $z_1=Z_1(\bar{q})$, $z_2=Z_2(\bar{q})$, and $z_m=M$. This equilibrium is described diagrammatically in Figure 6.

The curve *TT* describes the economy's transformation curve between the existing types of real equities (*TT* describes the *efficient* points). The distance *0A* equals *M*, the import quota. An equilibrium obtains at the point of tangency of an asset indifference surface, defined by all combinations of (z_1, z_2, z_m) which satisfy

$$(14) \quad Ev\{\bar{p}(\alpha);\, \theta_1(\alpha)z_1+\bar{p}(\alpha)\theta_2(\alpha)z_2 + [\bar{p}(\alpha)-p(\alpha)]z_m\} = \text{constant}$$

and *TT*. Suppose this happens at point E_s'. Then we can project the description of the equilibrium on the (z_1, z_2) plane, as described by the broken lines in Figure 6. This is identical to the equilibrium described in Figure 4 for the economy in which import quotas are auctioned off. The equilibrium relative real equity prices are

$$\bar{q}=\frac{Ev_I(\alpha)\bar{p}(\alpha)\theta_2(\alpha)}{Ev_I(\alpha)\theta_1(\alpha)}$$

$$q_m=\frac{Ev_I(\alpha)[\bar{p}(\alpha)-p(\alpha)]}{Ev_I(\alpha)\theta_1(\alpha)}$$

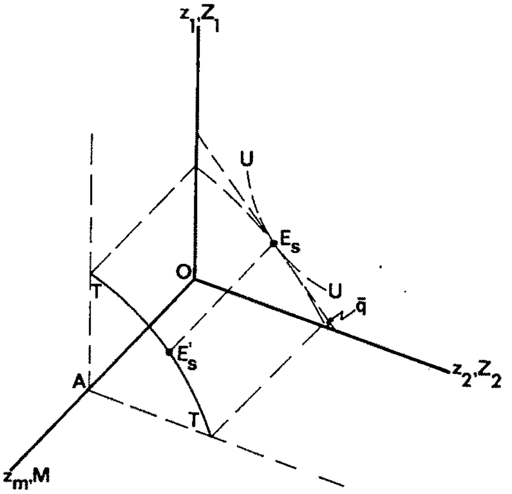


FIGURE 6

The reader can verify that these problems have identical solutions in terms of $\bar{q}$ and $(z_1, z_2)=(Z_1, Z_2)$ with $T(\alpha)=[\bar{p}(\alpha)-p(\alpha)]M=[\bar{p}(\alpha)-p(\alpha)][c_2(\alpha)-\theta_2(\alpha)Z_2]$.

III. The Most Preferred Protective Policy

We show in this section that a subsidy to domestic real equities of the import-competing sector, designed to achieve a desired level of protection of this industry, is preferred in terms of the resulting expected utility level (with one exception) to both a tariff and an import quota which achieve the same level of protection. The comparison with the import quota applies only to the case of no international trade in securities while the comparison with the tariff is independent of the existence of international trade in securities. In addition, we show that this policy dominates taxes on international financial capital flows when such flows do exist, as well as quotas on capital imports.

Suppose that we start with a free trade equilibrium in which the import-competing sector produces Z_2' units of real equities, and that we want to induce a resource flow into the import-competing sector so as to make it produce Z_2 units of real equities, $Z_2 > Z_2'$. Consider first the case of no international trade in securities, and suppose that

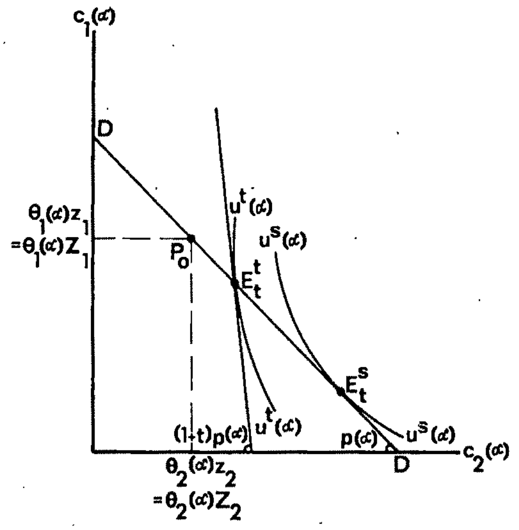


FIGURE 7

a subsidy at rate s on sales of type-2 real equities, a tariff at rate t, and a quota of size M *all* generate the same final production allocation $(Z_1, Z_2)=(z_1, z_2)$. The resulting equilibria in a particular state α under the three alternative policies are described in Figure 7. Point P_o, which represents domestic output levels in state α that result from the policy-induced allocation of productive resources, describes the consumption point achievable by the economy without international trade in commodities. If the economy trades in commodities it can choose a point on DD, whose slope is $p(\alpha)$, the foreign relative price of good 2. Under the subsidy, the economy chooses to consume at E_t^s and obtains in state α the utility level indicated by the indifference curve $u^s(\alpha)u^s(\alpha)$. Under the tariff, the economy chooses to consume at E_t^t, the point on DD at which the slope of the indifference curve that passes through it, $u^t(\alpha)u^t(\alpha)$, is equal to $(1+t)p(\alpha)$. Clearly, the welfare level under the tariff is smaller than the welfare level under the subsidy. Since this is true for every state α, the expected utility level achieved with the subsidy exceeds the expected utility level achieved by the tariff.

If a quota is used, there are two possibilities. First, the quota may not be binding in state α. In this case, the economy chooses to consume at E_t^s and its welfare level under the quota is the same in this state as under the subsidy. Second, the quota may be binding in state α. In this case, the economy consumes at a point on DD to the left of E_t^s, and its welfare level in state α is lower under the quota than under the subsidy. Since the quota has to be binding in at least some states of nature in order to achieve the desired level of protection, the expected utility level under the quota is necessarily lower than under the subsidy. Hence, the subsidy on equity sales is preferable to both the tariff and the quota.[5]

Now consider the case in which there is international trade in securities. If the tariff is to achieve the same level of protection as the subsidy, that is, if the production point on TT is to be the same in both cases, the tariff rate has to equal the rate of subsidy, $t=s$, which assures the same relative real equity prices faced by firms. These relative real equity prices are equated to the marginal rate of transformation in the production of real equities, thus determining the same allocation of productive resources.

In what follows, we prove that the tariff provides a lower expected utility level than the equity subsidy, except for one case. This proof can be easily extended to the case in which there exist foreign securities with distributions of returns that differ from the distribution of returns on domestic real equities. Using superscripts t and s to denote variables which apply to equilibria under the tariff and subsidy, respectively, we have from the concavity of the ordinary utility function $u(\cdot)$:

---

[5] Our discussion of the quota should be qualified. The above discussion is valid if indeed a quota assures protection of the import-competing sector. However, we do not have a proof of this proposition for the general case. If a tariff is not protective, we can use an export subsidy in order to achieve a desired level of protection and use the above techniques to prove the superiority of an equity subsidy on the export subsidy. But if a quota is not protective, we need to impose a minimal required level of exports or imports in order to achieve the desired protection, and this is difficult to enforce. Finally, observe that instead of using import quotas, one may use export quotas; the two are not equivalent in the presence of price uncertainty. However, qualitatively, everything we say about import quotas applies to export quotas as well.

$$(15)\quad \sum_{\alpha=1}^{S} \pi(\alpha)\left[u^t(\alpha)-u^s(\alpha)\right] < \sum_{\alpha=1}^{S} \pi(\alpha)\nabla u^s(\alpha)\left\{\left[c_1^t(\alpha),c_2^t(\alpha)\right] - \left[c_1^s(\alpha),c_2^s(\alpha)\right]\right\}$$

where $\pi(\alpha)$ is the subjective probability of state α and $\nabla u(\alpha)$ is the gradient of the function $u(\cdot)$ in state α. From the solution to (3)

$$(16)\quad \nabla u^s(\alpha)=v_I^s(\alpha)\left[1,p(\alpha)\right]$$

where $v_I^s(\alpha)$ is the marginal utility of income (in terms of good 1) in state α. Combining (16) with (15), we obtain

$$(17)\quad \sum_{\alpha=1}^{S} \pi(\alpha)\left[u^t(\alpha)-u^s(\alpha)\right] < \sum_{\alpha=1}^{S} \pi(\alpha)v_I^s(\alpha) \cdot\left[c_1^t(\alpha)+p(\alpha)c_2^t(\alpha)-c_1^s(\alpha)-p(\alpha)c_2^s(\alpha)\right]$$

From the budget constraint in (3)

$$(18)\quad c_1^t(\alpha)+(1+t)p(\alpha)c_2^t(\alpha)=\theta_1(\alpha)z_1^t+(1+t) \cdot p(\alpha)\theta_2(\alpha)z_2^t+tp(\alpha)\left[c_2^t(\alpha)-\theta_2(\alpha)Z_2^t\right]$$

$$(19)\quad c_1^s(\alpha)+p(\alpha)c_2^s(\alpha)=\theta_1(\alpha)z_1^s+p(\alpha)\theta_2(\alpha)z_2^s$$

Combining (17) with (18) and (19), we obtain

$$(20)\quad \sum_{\alpha=1}^{S} \pi(\alpha)\left[u^t(\alpha)-u^s(\alpha)\right] < \sum_{\alpha=1}^{S} \pi(\alpha)v_I^s(\alpha) \cdot\left[\theta_1(\alpha)(z_1^t-z_1^s)+p(\alpha)\theta_2(\alpha)(z_2^t-z_2^s) +tp(\alpha)\theta_2(\alpha)(z_2^t-Z_2^t)\right]$$

From the budget constraint in (4)

$$(21)\quad z_1^t+(1+t)qz_2^t=Z_1^t+(1+t)qZ_2^t$$

$$(22)\quad z_1^s+qz_2^s=Z_1^s+(1+s)qZ_2^s-sqZ_2^s$$

However, since the tariff and the subsidy provide the same level of protection,

$$(23)\quad Z_j^t=Z_j^s=Z_j \qquad j=1,2$$

$$(24)\quad t=s$$

Combining (20) with (21)–(24), we obtain

$$(25)\quad \sum_{\alpha=1}^{S} \pi(\alpha)\left[u^t(\alpha)-u^s(\alpha)\right] < \left[(1+t)(Z_2-z_2^t)-(Z_2-z_2^s)\right] \cdot \sum_{\alpha=1}^{S} \pi(\alpha)v_I^s(\alpha)\left[\theta_1(\alpha)q-p(\alpha)\theta_2(\alpha)\right]$$

Now, in the presence of an interior solution to (4) for the subsidy case,

$$(26)\quad \sum_{\alpha=1}^{S} \pi(\alpha)v_I^s(\alpha)\left[\theta_1(\alpha)q-p(\alpha)\theta_2(\alpha)\right]=0$$

which implies, using (25),

$$(27)\quad \sum_{\alpha=1}^{S} \pi(\alpha)u^t(\alpha) < \sum_{\alpha=1}^{S} \pi(\alpha)u^s(\alpha)$$

Hence, the subsidy provides a higher expected utility level than the tariff whenever the investor does not specialize in his portfolio under the subsidy system. He may want to specialize due to the prohibition of short sales.[6] If he chooses to specialize in type-2 real equities, the expression on the left-hand side of (26) is nonpositive. Then

---

[6] To reformulate problem (4) to allow for short sales, we have to add nonbankruptcy constraints. Then, if the solution to the new problem is such that the bankruptcy constraints are not binding in all states of nature, condition (26) is satisfied and the equity subsidy dominates the tariff.

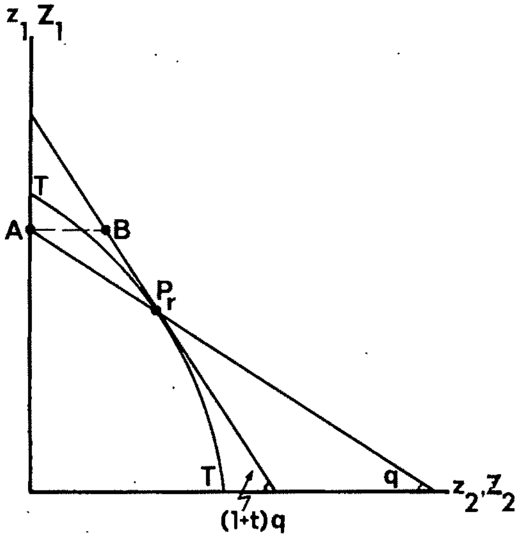


FIGURE 8

also $Z_2 - z_2^s = -Z_1/q$, while the smallest possible value of $Z_2 - z_2^t$ is $-Z_1/[(1+t)q]$ (it is obtained by choosing the largest feasible value of z_2^t, which obtains when $z_1^t = 0$). Hence

$$(1+t)(Z_2 - z_2^t) - (Z_2 - z_2^s) \leq 0$$

These considerations show that (27) is satisfied in the case in which the investor chooses to hold only type-2 real equities, thus proving that the equity subsidy dominates the tariff also in this case.

If, as a result of restrictions on short sales, the investor chooses under the subsidy to specialize in his portfolio in type-1 (the exporting sector's) real equities, it may happen that the tariff will be preferred to an equity subsidy. (There is no reason to have identical patterns of trade in goods and securities.) The reason for this is the fact that in the presence of ("strong") preferences for type-1 real equities the short-sales restriction may be more binding in the subsidy regime than in the tariff regime. This is demonstrated in Figure 8.

Point P_r represents the production point under the tariff and subsidy. Under the subsidy the investor chooses, by assumption, point A. The choice of point A implies that the left-hand side of (26) is nonnegative. Now, if under the tariff the investor chooses a point to the right of B on the line with slope $(1+t)q$, $(1+t)(Z_2 - z_2^t) - (Z_2 - z_2^s) > 0$, and (25) implies (27). In this case, the equity subsidy is preferred to the tariff. But it may happen that the portfolio point under the tariff is to the left of B on the line with slope $(1+t)q$. In this case the welfare losses from the short-sales restriction which are operative in the subsidy regime, may outweigh the welfare losses from tariffs. A similar argument is used in our forthcoming paper to show that in the presence of international trade in securities a *small* country may *gain* from a tariff. The following is an example in which the tariff dominates the equity subsidy.

Example: Let $u = c_1 + \log c_2$, then $c_1(\alpha) = \theta_1(\alpha)z_1 + \bar{p}(\alpha)\theta_2(\alpha)z_2 + T(\alpha) - 1$, $c_2(\alpha) = 1/\bar{p}(\alpha)$, and $v(\alpha) = \theta_1(\alpha)z_1 - \bar{p}(\alpha)\theta_2(\alpha)z_2 + T(\alpha) - 1 - \log \bar{p}(\alpha)$. Assume that $E\theta_1(\alpha) = 1$, $Ep(\alpha)\theta_2(\alpha) < q$, and $1/p(\alpha) - \theta_2(\alpha)Z_2 > 0$ for all α, where Z_2 is at the desired level of protection. It is now easy to see that the solution to (4) is such that the investor specializes in z_1 in both the tariff and equity subsidy regimes. This occurs because his asset indifference curves are straight lines with slope $E\bar{p}(\alpha)\theta_2(\alpha)$, where $\bar{p}^t(\alpha) = (1+t)p(\alpha)$ and $\bar{p}^s(\alpha) = p(\alpha)$, which is smaller than q_c in either case, since $q_c^t = (1+t)q$ and $q_c^s = q$. This means that

$$z_1^t = Z_1 + (1+t)qZ_2,\ z_2^t = 0,$$

$$z_1^s = Z_1 + qZ_2,\ z_2^s = 0$$

$$T^t(\alpha) = t/(1+t) - tp(\alpha)\theta_2(\alpha)Z_2,$$

$$\alpha = 1, 2, \ldots, S,\ T^s(\alpha) = 0$$

Hence

$$Ev^t(\alpha) - Ev^s(\alpha) = tZ_2[q - Ep(\alpha)\theta_2(\alpha)] + t/(1+t) - \log(1+t)$$

which is positive for sufficiently small values of t and $Z_2[q - Ep(\alpha)\theta_2(\alpha)]$ sufficiently large.

Let us compare the results of this section to similar results that obtain in deterministic models. It is well-known that in a deterministic world a subsidy to the output of the import-competing sector is a better means of protection than a tariff. The reason is that with the same composition of production in the presence of a tariff there is a distortion which results from the difference between consumer prices and foreign prices. This distortion is absent under the output subsidy. The reader can verify that a similar argument applies under uncertainty to the comparison of a tariff with a subsidy on equity sales of the import-competing sector. There are however other considerations which have to be taken into account in the presence of international trade in securities, for example, restrictions on short sales of securities and absence of bankruptcies. In order to make the analogy complete, observe that a fully anticipated subsidy to output sales of the import-competing sector after the realization of state of nature, the subsidy being financed by means of *ex post* lump sum taxes, is equivalent to an equity subsidy at the same rate. Both increase the producer-relative price of the import-competing sector's equities above the investor's price at the same rate, and make consumption decisions depend on foreign commodity prices. However, unlike a deterministic world, a subsidy to commodity sales (or to equity sales) is not always preferable to a tariff, as shown in this section.

IV. Tariffs vs. Quotas in the Absence of International Trade in Securities

The discussion in this section is restricted to the case in which there is no international trade in securities, since (as explained in Section II) there is no satisfactory way to deal in a single country model with quotas in the presence of international trade in securities. We showed in the previous section that in the case considered here an equity subsidy is preferred to an ad valorem tariff and to a fixed quota as protective policy measures. Now we confine the attention to a comparison of welfare levels which result from a tariff and a quota designed to achieve the same level of protection for the import-competing industry. Our conclusion is that there are circumstances in which a tariff is preferred to a quota and vice versa, and we demonstrate it as follows.

Example: Let the utility function be $u=f(c_1+\log c_2)$. Then the demand functions are $c_1=I-1$ and $c_2=1/\bar{p}$, and the indirect utility function is $v=f(I-1-\log\bar{p})$.

Suppose that we start from a position of full certainty. Then tariffs and quotas are equivalent means of protection. Let t_o be the tariff rate that provides the desired level of protection with (Z_1, Z_2) being the levels of output of real equities. Now suppose that we introduce a "little" uncertainty adjusting the tariff rate and the quota so as to maintain (Z_1, Z_2) as the equilibrium allocation. By developing a second-order approximation to the loss of expected utility under the tariff and quota, we can compare the welfare losses and see when the tariff is preferred to the quota and vice versa.

Let

$$\text{(28a)}\quad \theta_1(\alpha,h)\equiv 1+h\varepsilon(\alpha)\qquad \alpha=1,2,\ldots,S$$

$$\text{(28b)}\quad \theta_2(\alpha)\equiv 1\qquad \alpha=1,2,\ldots,S$$

$$\text{(29)}\quad p(\alpha,h)\equiv 1+h\delta(\alpha)\qquad \alpha=1,2,\ldots,S$$

where $E\varepsilon(\alpha)=E\delta(\alpha)=0$, $E[\varepsilon(\alpha)]^2=\sigma_\varepsilon^2$, $E[\delta(\alpha)]^2=\sigma_\delta^2$, $E\varepsilon(\alpha)\delta(\alpha)=\sigma_{\varepsilon\delta}^2$. For $h=0$ we have full certainty. Hence, if $\bar{p}(\alpha,h)$ stands for the domestic relative price of good 2 in state α,

$$\text{(30)}\quad \bar{p}(\alpha,0)=1+t_o=MRT\qquad \alpha=1,2,\ldots,S$$

where MRT is the marginal rate of transformation along the real equity transformation curve at point (Z_1, Z_2), and (30) holds for both the tariff and the quota. From the maximization problem shown as (4), equilibrium requires

$$\text{(31)}\quad Ev_I(\alpha,h)\bar{p}(\alpha,h) - MRT\,Ev_I(\alpha,h)\theta_1(\alpha,h)=0$$

We represent a "small" increase in uncertainty by a shift of h from zero to a small positive h.

Let

$$(32)\quad U(h) \equiv Ev\{\bar{p}(\alpha,h);\theta_1(\alpha,h)Z_1 + \bar{p}(\alpha,h)Z_2 + [\bar{p}(\alpha,h) - p(\alpha,h)]M(\alpha,h)\}$$

be the equilibrium expected-utility level, expressed as a function of h, where $M(\alpha, h)$ is equal to state α imports given by

$$(33)\quad M(\alpha,h) = \bar{p}(\alpha,h)^{-1} - Z_2$$

Then

$$(34)\quad U(h) = U(0) + U'(0)h + (1/2)U''(0)h^2 + \textit{residual}$$

Clearly, $U(0)$ is the same for a quota and a tariff. We have to calculate $U'(0)$ and $U''(0)$ for a quota and a tariff. These calculations are presented in the Appendix. They yield $U'(0)=0$ for a tariff and a quota. Hence, the differences arise only from second- or higher-order terms. It is shown in the Appendix that

$$(35)\quad U''(0)^{tariff} = v_I(0) \cdot \left\{\sigma_\delta^2 - 2A\frac{t_o}{1+t_o}[(1-Z_2+Z_1)\sigma_{\epsilon\delta}^2 - (1-Z_2)\sigma_\delta^2 - Z_1\sigma_\epsilon^2] - A[Z_1^2\sigma_\epsilon^2 + (1-Z_2)^2\sigma_\delta^2 - 2Z_1(1-Z_2)\sigma_{\epsilon\delta}^2]\right\}$$

$$(36)\quad U''(0)^{quota} = v_I(0)\left\{2A\frac{t_o}{1+t_o}[Z_1\sigma_\epsilon^2 - M(0)\sigma_{\epsilon\delta}^2] - A[Z_1^2\sigma_\epsilon^2 + M(0)^2\sigma_\delta^2 - 2Z_1M(0)\sigma_{\epsilon\delta}^2]\right\}$$

where A is the absolute measure of risk aversion, which depends on the concavity of $f(\cdot)$. Using (30) and (33)

$$(37)\quad M(0) = (1+t_o)^{-1} - Z_2$$

Now, using (34)–(37), we have

$$(38)\quad U(h)^{quota} - U(h)^{tariff} \cong v_I(0)\left[-\sigma_\delta^2 + A\frac{t_o^2}{(1+t_o)^2}(2\sigma_{\epsilon\delta}^2 - \sigma_\delta^2)\right]$$

Now observe that in the presence of risk neutrality, i.e., $A=0$, or when $2\sigma_{\epsilon\delta}^2 < \sigma_\delta^2$, the tariff dominates the quota. But if $2\sigma_{\epsilon\delta}^2 > \sigma_\delta^2$, and the degree of risk aversion is sufficiently large (large A), the quota dominates the tariff. This shows that a quota may be preferred to a tariff, and vice versa, depending on the nature of the economy.

V. Concluding Comments

We have shown that in order to achieve a desired level of protection of the import-competing sector in the absence of international trade in securities, an equity subsidy is preferred to both an ad valorem tariff with a predetermined rate and to a quota with a predetermined size. We also have shown that in this case a tariff may be preferred to a quota and vice versa, depending on the structure of risk and attitudes towards risk.

In the presence of international trade in securities, an equity subsidy dominates an ad valorem tariff whenever—under the equity subsidy—investors do not choose to specialize in holdings of assets of the domestic export sector and their perfect foreign substitutes. If domestic investors specialize in holdings of the exporting industry's equities as a result of restrictions on short sales, a tariff may dominate an equity subsidy.

Our analysis concentrated on policy instruments whose value is chosen *ex ante*. Thus, the tariff rate is chosen before the realization of a state of nature and the same rate applies to all states of nature; the quota is chosen before the realization of a state of nature and the same size quota applies to all states of nature (although in some states it may not be binding). The equity subsidy is chosen and applied at the stage of investment—before the realization of a state of nature. Observe, however, that (as explained in Section III) a subsidy to sales of output of the industry whose equities we wish to subsidize, at a rate equal to the rate of equity subsidy, achieves the same result as the equity subsidy if applied uniformly to all states of nature. Alternatively, the same result can be achieved by a subsidy at a uniform rate (equal to $s/(1+s)$, where s is the rate of equity subsidy) to employment of factors of production in the protected industry. Observe also that Abba Lerner's symmetry theorem applies in the present framework to ad valorem tariff and export tax rates. Thus, our uniform tariff rate can be replaced by an ad valorem tax on exports at a uniform rate (equal to $t/(1+t)$, where t is the tariff rate), bringing about the same allocation of resources. However, a specific tariff which is state independent does not have a state-independent specific export tax equivalent. The same is true for import and export quotas which are state independent —they are not equivalent.

The nonequivalence of the above mentioned instruments results from the determination of their level before a realization of a state of nature, and the application of these levels uniformly to every state of nature. If we are free to apply trade taxes and quotas at different rates and levels in different states of nature, that is, if we can choose the levels of these policy measures contingent on states of nature, then tariffs (specific and ad valorem), export taxes (specific and ad valorem), import quotas, and export quotas are all equivalent policies. For the state-dependent structure of any one of them there exist state-dependent structures of the other three which result in the same allocation of resources. We do, however, believe that state-contingent policies are difficult to execute, and that in many cases feasibility considerations necessitate the choice of state-independent levels of policy instruments.

Nevertheless, if state-dependent intervention is feasible and there is no international trade in securities, *there is no state-dependent policy which is strictly preferred to a subsidy on equity sales of the protected sector in order to achieve a desired level of protection.* This can be seen as follows. Given (Z_1, Z_2), in equilibrium $(z_1, z_2)=(Z_1, Z_2)$ and the expected utility level under any policy is

$$Ev\{\bar{p}^c(\alpha);\theta_1(\alpha)Z_1+\bar{p}^p(\alpha)\theta_2(\alpha)Z_2 + [\bar{p}^c(\alpha)-p(\alpha)]c_2(\alpha) - [\bar{p}^p(\alpha)-p(\alpha)]\theta_2(\alpha)Z_2\}$$

where $\bar{p}^c(\alpha)=$ the consumer-relative price in state α; $\bar{p}^p(\alpha)=$ the producer-relative price in state α; $c_2(\alpha)=$ equilibrium consumption of good 2 in state α; all induced by the policy under consideration. For example, under an *ex post* ad valorem tariff $\bar{p}^c(\alpha)=\bar{p}^p(\alpha)=[1+t(\alpha)]p(\alpha)$, while under an ad valorem uniform subsidy on the second sector's output $\bar{p}^c(\alpha)=p(\alpha)$ and $\bar{p}^p(\alpha)=[1+s]p(\alpha)$. Now, from a standard free trade argument:

$$v[p(\alpha);\theta_1(\alpha)Z_1+p(\alpha)\theta_2(\alpha)Z_2] \geqslant v\{\bar{p}^c(\alpha);\theta_1(\alpha)Z_1+\bar{p}^p(\alpha)\theta_2(\alpha)Z_2 + [\bar{p}^c(\alpha)-p(\alpha)]c_2(\alpha) - [\bar{p}^p(\alpha)-p(\alpha)]\theta_2(\alpha)Z_2\} \tag{39}$$

for any $\bar{P}^c(\alpha)$, $\bar{p}^p(\alpha)$, $\alpha=1,2,\ldots,S$. Since for a subsidy on equity 2, $\bar{p}^c(\alpha)=\bar{p}^p(\alpha)=p(\alpha)$, it is the best policy. Observe also that a state-independent output subsidy on good 2, financed by lump sum taxes, is also the best policy, since in this case $\bar{p}^c(\alpha)=p(\alpha)$ and $\bar{p}^p(\alpha)=(1+s)p(\alpha)$, so that for this policy the right-hand side of (39) is just equal to its left-hand side.

Finally, let us comment on the use of state-independent trade taxes and quotas when an equity subsidy is not available. In this case there is room for independent uses of tariffs, export taxes, import quotas, and export quotas (if we use ad valorem trade taxes, a tariff and export tax are equivalent).

Each one of these tools may perform a useful task and one may be interested in their optimal combination. However, this question is beyond the scope of this paper.

APPENDIX

Generally, from characteristics of indirect utility functions, (32) yields, using (28a) and (29),

(A1)

$$U'(h)=Ev_I(\alpha,h)\{\epsilon(\alpha)Z_1-\delta(\alpha)M(\alpha,h)+[\bar{p}(\alpha,h)-p(\alpha,h)]M'(\alpha,h)\}$$

where $M'(\alpha,h)=dM(\alpha,h)/dh$. From (A1),

$$\text{(A2)}\quad U''(h)=Ev_I(\alpha,h)\Big\{-\delta(\alpha)M'(\alpha,h)+\Big[\frac{d\bar{p}(\alpha,h)}{dh}-\delta(\alpha)\Big]M'(\alpha,h)+[\bar{p}(\alpha,h)-p(\alpha,h)]M''(\alpha,h)\Big\}+E\frac{dv_I(\alpha,h)}{dh}\{\epsilon(\alpha)Z_1-\delta(\alpha)M(\alpha,h)+[\bar{p}(\alpha,h)-p(\alpha,h)]M'(\alpha,h)\}$$

where

$$\text{(A3)}\quad \frac{dv_I(\alpha,h)}{dh}=-v_I(\alpha,h)c_{2I}(\alpha,h)\cdot\frac{d\bar{p}(\alpha,h)}{dh}+v_{II}(\alpha,h)\{\epsilon(\alpha)Z_1-\delta(\alpha)M(\alpha,h)+[\bar{p}(\alpha,h)-p(\alpha,h)]M'(\alpha,h)\}$$

However, $c_{2I}\equiv 0$, so that the first term can be disregarded. Differentiating the equilibrium condition (31), we get

(A4)

$$E\frac{dv_I(\alpha,h)}{dh}[\bar{p}(\alpha,h)-MRT\theta_1(\alpha,h)]+Ev_I(\alpha,h)\Big[\frac{d\bar{p}(\alpha,h)}{dh}-MRT\epsilon(\alpha)\Big]=0$$

Evaluating (A4) at $h=0$, using (28a) and (30) and remembering that $v_I(\alpha,0)=v_I(0)$ is state independent, we get

$$\text{(A5)}\quad E\frac{d\bar{p}(\alpha,0)}{dh}=0$$

for a tariff and a quota. Differentiating (A4), and evaluating it at $h=0$, we get

$$\text{(A6)}\quad 2Ev_{II}(0)[\epsilon(\alpha)Z_1-\delta(\alpha)M(\alpha,0)+t_oM'(\alpha,0)]\Big[\frac{d\bar{p}(\alpha,0)}{dh}-MRT\epsilon(\alpha)\Big]+Ev_I(0)\frac{d^2\bar{p}(\alpha,0)}{dh^2}=0$$

A. *Tariff*

Consider now the ad valorem tariff. If $t(h)$ is the equilibrium tariff rate,

$$\text{(A7)}\quad \bar{p}(\alpha,h)=[1+t(h)][1+h\delta(\alpha)]\qquad \alpha=1,2,\ldots,S$$

Hence

$$\text{(A8)}\quad \frac{d\bar{p}(\alpha,h)}{dh}=t'(h)[1+h\delta(\alpha)]+[1+t(h)]\delta(\alpha)$$

$$\text{(A9)}\quad \frac{d^2\bar{p}(\alpha,h)}{dh^2}=t''(h)[1+h\delta(\alpha)]+2t'(h)\delta(\alpha)$$

Combining (A8) with (A5) implies $t'(0)=0$. Hence,

$$\text{(A10)}\quad \frac{d\bar{p}(\alpha,0)}{dh}=(1+t_o)\delta(\alpha)\qquad \alpha=1,2,\ldots,S$$

$$\text{(A11)}\quad \frac{d^2\bar{p}(\alpha,0)}{dh^2}=t''(0)\qquad \alpha=1,2,\ldots,S$$

From (33), using (A10) and (A11),

$$\text{(A12)}\quad M'(\alpha,0)=-\delta(\alpha)/(1+t_o)\qquad \alpha=1,2,\ldots,S$$

$$\text{(A13)}\quad M''(\alpha,0)=-t''(0)/(1+t_o)^2 +2[\delta(\alpha)]^2/(1+t_o) \qquad \alpha=1,2,\dots,S$$

From (A6), using (A10)–(A12), we solve

$$\text{(A14)}\quad t''(0)=2A(1+t_o)\left[(1-Z_2+Z_1)\sigma_{\epsilon\delta}^2 - (1-Z_2)\sigma_\delta^2-Z_1\sigma_\epsilon^2\right]$$

where $A=-v_{II}(0)/v_I(0)=-f''(0)/f'(0)$ is the measure of absolute risk aversion at the initial, nonstochastic equilibrium point. Using (A1) and (A12), we get

$$\text{(A15)}\qquad U'(0)^{tariff}=0$$

and using (30), (A2), (A3), (A10), (A12), and (A13), we get

$$\text{(A16)}\quad U''(0)^{tariff}$$

$$=v_I(0)\left\{\sigma_\delta^2-2A\frac{t_o}{1+t_o}\left[(1-Z_2+Z_1)\sigma_{\epsilon\delta}^2 -(1-Z_2)\sigma_\delta^2-Z_1\sigma_\epsilon^2\right] -A\left[Z_1^2\sigma_\epsilon^2+(1-Z_2)^2\sigma_\delta^2-2Z_1(1-Z_2)\sigma_{\epsilon\delta}^2\right]\right\}$$

B. *Quota*

From (33), for a state-independent quota $M(h)$ which is binding in every state, as it is in our case for small h (since initially the quota was strictly binding), we have

$$\text{(A17)}\quad \bar{p}(\alpha,h)=[Z_2+M(h)]^{-1} \qquad \alpha=1,2,\dots,S$$

Hence, $\bar{p}(\alpha,h)=\bar{p}(h)$ is state independent, and using (A5),

$$\text{(A18)}\qquad d\bar{p}(0)/dh=0$$

Now (A17) and (A18) imply

$$\text{(A19)}\qquad M'(0)=0$$

Using (A18) and (A19), we solve from (A6):

$$\text{(A20)}\quad \frac{d^2\bar{p}(0)}{dh^2}=-2A(1+t_o)\left[Z_1\sigma_\epsilon^2-M(0)\sigma_{\epsilon\delta}^2\right]$$

Combining (A29) with (A17),

$$\text{(A21)}\quad M''(0)=2A\frac{1}{1+t_o}\left[Z_1\sigma_\epsilon^2-M(0)\sigma_{\epsilon\delta}^2\right]$$

From (30), (A1), and (A19):

$$\text{(A22)}\qquad U'(0)^{quota}=0$$

and from (30), (A1), (A2), (A18), (A19), and (A21):

$$\text{(A23)}\quad U''(0)^{quota}$$

$$=v_I(0)\left\{2A\frac{t_o}{1+t_o}\left[Z_1\sigma_\epsilon^2-M(0)\sigma_{\epsilon\delta}^2\right] -A\left[Z_1^2\sigma_\epsilon^2+M(0)^2\sigma_\delta^2-2Z_1M(0)\sigma_{\epsilon\delta}^2\right]\right\}$$

REFERENCES

P. Dasgupta and J. E. Stiglitz, "Tariffs vs. Quotas as Revenue Raising Devices under Uncertainty," *Amer. Econ. Rev.*, Dec. 1977, *67*, 975–81.

G. Fishelson and F. Flatters, "The (Non)-Equivalence of Optimal Tariffs and Quotas Under Uncertainty," *J. Int. Econ.*, Nov. 1975, *5*, 385–93.

E. Helpman and A. Razin, (1978a) "Welfare Aspects of International Trade in Goods and Securities," *Quart. J. Econ.*, Aug. 1978, *92*, 489–508.

______**and**______, (1978b) "The Protective Effect of a Tariff Under Uncertainty," *J. Polit. Econ.*, Dec. 1978, *86*, 1131–42.

______**and**______, (1978c) *A Theory of International Trade Under Uncertainty*, New York 1978.

______**and**______, "Welfare Aspects of International Trade in Goods and Securities: An Addendum," *Quart. J. Econ.*, forthcoming.

Unemployment, Inflation, and Monetarism— A Further Analysis: Comment

By J. KIRKER STEPHENS*

In a recent paper in this *Review*, Robert Van Order proposes an interesting simplification of the important model of Jerome Stein (1974). The purpose of this comment is to correct some errors in Van Order's analysis, to relate his results to Stein's later work (1976b,c), and to analyze the role of these papers in the current debate over the mechanisms and effectiveness of macroeconomic policy.

For the convenience of the reader and to simplify the necessary references to Van Order's paper, his equations through (7′) and relevant notation are reproduced:

g, g^*: actual and expected rates of inflation
y: real output
f: indicator of fiscal policy
M, m: nominal and real money supply
P, W: levels of prices and wages
V: velocity of circulation
n: growth rate of M
λ: speed of adjustment of expected inflation
θ: ratio of bonds to money

Functions:

P: Phillips curve
F: production function
E: aggregate demand
S: aggregate supply
R: reduced form of E and S

Van Order's equations are as follows:

$$(1) \quad g = P(y) + g^* \qquad P' > 0$$

$$(2) \quad Dg^* = \lambda(g - g^*) \qquad \lambda > 0$$

$$(3) \quad Dg^* = \lambda P(y)$$

$$(4) \quad P/W = S(y)$$

$$(5) \quad P/W = (M/W) \cdot E(y, f, g^*, \theta)$$

$$(5') \quad Py/M = yE(y, f, g^*, \theta) = V$$

$$(6) \quad y = R(f, m, \theta, g^*) \quad \text{with } m = M/W$$

$$(7) \quad Dy = R_f Df + R_m Dm + R_\theta D\theta + R_{g^*} Dg^*$$

$$(7') \quad Dm = m(n - g)$$

I. Correction of Van Order's Analysis

To derive equation (8) Van Order says, "Assume f, n, and θ are constant. Then from (1) and (2), (7) becomes

$$(8) \quad Dy = -R_m m[P(y) + g^*] + R_{g^*} \lambda P(y)\text{"}$$

(p. 743)

This, however, is incorrect. Assuming that n is constant, but not necessarily equal to zero, correct algebra yields[1]

$$Dy = -R_m m[P(y) + g^*] + R_{g^*} \lambda P(y) + R_m mn$$

Appropriate related corrections must be made to his succeeding equations also.

Indeed, the correction of equation (8) is necessary for consistency with Van Order's phase diagram for the system, which he says is derived from equations (3) and (8). (See my Figure 1 which is essentially his Figure 1 with some additions.) If (3) and (8) hold, then the long-run equilibrium value of g^* would have to be 0, instead of n as in Figure 1, since the value of y that made the right-hand side of (3) equal to 0 would make the right-hand side of (8) equal to $-R_m mg^*$, which (assuming $R_m \neq 0$) can only be 0 when $g^*=0$. Thus the conditions $Dg^*=0$ and $Dy=0$ could only be fulfilled simultaneously when $g^*=0$.

---

*Professor of economics, University of Oklahoma.

[1] Van Order later states, "In the long run the time derivatives are zero..." (p. 743). One might then possibly wonder whether he meant to assume M, rather than n, constant, in which case equation (8) would be correct since n would then be 0. However, as will be shown in the text, this is inconsistent with his Figure 1, in which the equilibrium value of g^* is not 0. Thus the statement that the time derivatives equal zero must be interpreted as referring only to those appearing in equation (7).

Van Order later says, "Consider a decline in the rate of monetary growth. This shifts the $Dg^*=0$ line downward, leaving the long-run level of output unchanged" (p. 744). This is surely a misprint; it is the $Dy=0$ rather than the $Dg^*=0$ line that is shifted downward. With that correction the analysis is reasonable.

He then says, "The simplified model brings out the issue of 'stagflation,' i.e., high inflation and high unemployment at the same time. Suppose that f is increased, raising aggregate demand. This raises both employment and inflation. Eventually we might expect f to be cut back perhaps to its initial level, putting us in a position like that at point D in the figure" (p. 744). This analysis is at best confused or misleading,[2] and in any event is not the most interesting case.

The most interesting case, because of its role in the Keynesian-Monetarist debate (see, for example, Stein, 1976b, and Robert Solow), is the case where the fiscal variable f is increased and then maintained at its new higher level. Equation (7) does not contain f, but it does contain Df. Thus while f is *increasing*, the $Dy=0$ curve is shifted upward, say to the position $Dy=0$ ($Df>0$) in Figure 1. Output and inflation would initially rise, although the system would be headed for an equilibrium at point E with output back at its initial level but higher inflation. However, if f is then maintained at a constant increased *level*, Df will again equal zero and the $Dy=0$ curve will return to its original position. Suppose that, at the time the $Dy=0$ curve returned to its original position, the economy had reached point D. Then the economy would follow the path DCO back to the original (and now reestablished) equilibrium point. Thus the path Van Order describes is appropriate to the case of a permanently maintained increase in the fiscal variable; it does not require that f be cut back to its initial level.

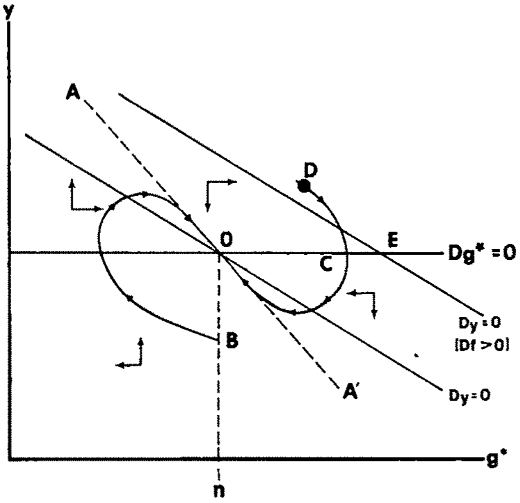


FIGURE 1. TIME PATHS OF OUTPUT AND EXPECTED INFLATION

II. Relations to Stein's Later Work

On properly analyzing the Van Order model, its long-run result when f is increased to a maintained higher level is that the rate of (wage) inflation g and the level of output y are unchanged. These results concerning the effect of a maintained change in fiscal policy which does not affect the long-run rate of inflation via changing θ are entirely consistent with those obtained by Stein.

---

[2] Van Order refers to an increase in f, followed by its return to its original level. He appears to believe this will shift the $Dy=0$ line up and then back to its original position. Actually when $Df>0$, the $Dy=0$ curve will be above its original position (say at Y_1Y_1' in Figure 2). While f is maintained at a higher level Df will equal 0 and the curve will return to its original position (Y_0Y_0' in Figure 2). When f is decreased toward its original level, Df will be negative for a time so the $Dy=0$ curve will be below its original position for a time (say at Y_2Y_2' in Figure 2). Finally, when f stabilizes at its original level, the $Dy=0$ line will once again resume its original position. The path y and g^* will actually follow throughout this process will depend on the speed of adjustment of y and g^* compared to the timing of the shifts in the curve.

Van Order says, "Here we keep with the standard assumption (as in Stein's model) that M and θ are exogenous" (pp. 742–43). (As he acknowledges, because of the government budget constraint, changes in f in his model must therefore be of a "balanced budget" type.) However, Stein (1974) considers situations where θ is endogenous; for example, "A rise in θ (say, resulting from a tax reduction where the deficit is financed through bond sales)..." (p. 886). Stein (1976b) considers the effect of "nonbalanced-budget" fiscal policy on the economy via its effect on θ in greater detail.

Stein says, "Within the context of this [1976b] model, the entire monetarist-fiscalist (neo-Keynesian) controversy hinges upon the sign of P_5" (p. 194), (see also 1976c). In his model the sign of P_5 determines the direction of the long-run effect, if any, on the rate of inflation of a maintained deficit (or surplus) financed by issuing (or retiring) bonds. The sign of R_θ in fact plays virtually the same role in Van Order's model, although Van Order takes no note of this.

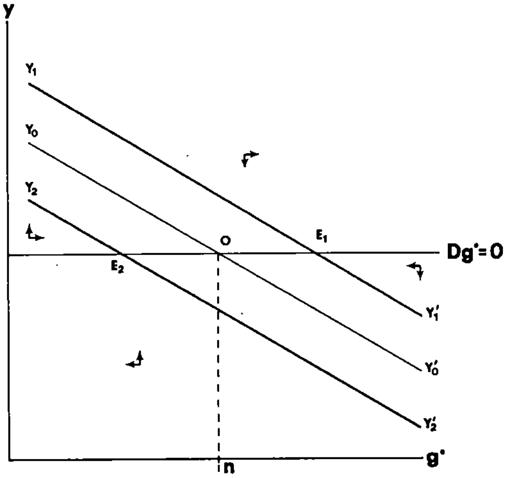


FIGURE 2. EQUILIBRIA OF OUTPUT AND EXPECTED INFLATION

Consider the situation where an increase in government purchases (of goods) to a maintained higher level is financed by bond sales. The long-run effect of this in Stein's model will be to increase the rate of (price) inflation if $P_5 > 0$ (his "Keynesian" case). The rate of inflation will remain unchanged if $P_5 = 0$, and will be lowered if $P_5 < 0$ (his "monetarist" cases). Now consider the same situation in Van Order's model, making θ endogenous to allow the possibility of such a nonbalanced-budget change in fiscal policy. Starting from an initial equilibrium with the $Dy = 0$ line at Y_0Y_0' in Figure 2, the $Dy = 0$ line will have an initial impetus shifting it up while $Df > 0$, and another impetus (possibly) shifting it in a direction determined by the sign of R_θ since $D\theta > 0$ (see equation (7)). Once f has reached its new maintained level the first effect will disappear and only that due to the term $R_\theta D\theta$ will remain. This effect (unless $R_\theta = 0$) will be permanent since the maintained deficit will require continual issuing of additional bonds to finance it. If $R_\theta > 0$ the $Dy = 0$ line will be permanently shifted up (say to Y_1Y_1'), and the new equilibrium E_1 will be one with a higher rate of (wage) inflation than at the initial equilibrium. If $R_\theta = 0$ then the $Dy = 0$ line will return to its initial position and the initial equilibrium at 0 will be reestablished. If $R_\theta < 0$ the $Dy = 0$ line will end up shifted down (say to Y_2Y_2') and the new equilibrium E_2 will be one with a lower rate of inflation than the initial equilibrium. Thus the sign of R_θ indeed plays virtually the same role in the long-run analysis of Van Order's model that the sign of P_5 plays in the long-run analysis of Stein's. And as Van Order indicates, the sign of R_θ is "...ambiguous since a rise in bond holdings tends to raise aggregate demand via a wealth effect and to lower it by increasing the demand for money..." (p. 742). In this regard, also, R_θ is analogous to P_5.

III. Output and Employment in the Long Run

Van Order says, "A Keynesian position might be that the $Dg^* = 0$ line is not flat, assuming that the coefficient of g^* in (1) is less than unity, but that is not of central importance" (p. 745). It is hard to see how this can fail to be of central importance since the question of whether the relation $Dg^* = 0$ is *horizontal* and *not shifted by fiscal or monetary policy* determines whether the long-run equilibrium level of output (and unemployment) can be affected by ordinary macro-economic policy.

There are at least three reasons why the relation $Dg^*=0$ may not be fixed and horizontal. First, as Van Order indicates, the coefficient of g^* in (1) might not equal one. Suppose his (1) is generalized as

$$g=P(y)+\alpha g^* \qquad P'>0, \alpha \leqslant 1 \tag{1a}$$

Then his (3) is replaced by

$$Dg^*=\lambda[P(y)-(1-\alpha)g^*] \tag{3a}$$

which gives the $Dg^*=0$ line a positive slope when $\alpha<1$.

Second, even if g^* has a coefficient of unity the function P may not be independent of g^*. Stein says of the unemployment term in his equation for wage change:

> The *h* function is a macro-economic relation which reflects the distribution of employment and unemployment among the various sectors of the economy. It is quite possible that the *h* function is related to the dispersion of the unemployment or to the composition of the labor force which, in turn, is related to the actual rate of price change π...even though the coefficient of the expected rate of price change...is unity. Therefore, the *h* function may be $h(U,\pi)$ instead of $h(U)$. [1976b, p. 187][3]

Since the rate of price change is related to the rate of wage change which, in turn, is related to the expected rate of wage change, the analogous situation in the Van Order model would be to replace his equation (1) by

$$g=P(y,g^*)+g^* \tag{1b}$$

so his (3) would be replaced by

$$Dg^*=\lambda P(y,g^*) \tag{3b}$$

which would also yield a nonhorizontal $Dg^*=0$ line.[4]

Third, if workers are primarily interested in after-tax real wages rather than before-tax real wages, a change in tax rates will change the *long-run equilibrium* level of employment and output. That is, by changing the differential between before-tax and after-tax real wages, it will change the level of employment (and so of y) at which workers are providing the amount of labor they wish to at the (after-tax) real wage they are receiving and employers are employing the amount of labor they wish to at the (before-tax) real wage they are paying.[5] Thus changes in f can change the equilibrium value of y so Van Order's (1) will be replaced by

$$g=P(y,f)+g^* \tag{1c}$$

Therefore his (3) would be replaced by

$$Dg^*=\lambda P(y,f) \tag{3c}$$

In this case the $Dg^*=0$ line will be horizontal for any given value of f, but can be permanently shifted by a change in f to a different maintained level.

Any of these three effects yields a result importantly different from that of Van Order's model, namely that at least some forms of fiscal or monetary policy can affect the long-run equilibrium level of output and employment.

IV. Conclusion

In considering the question of what issues constitute the differences between the Monetarists and the Keynesians, Van Order makes the following very useful suggestion: "It is convenient to separate these issues into long run, short run and medium run, the first two relating to comparative statics in either the long run or short run and the last relating to comparative dynamics" (p. 744). He then emphasizes the medium run, concluding "...the simple model enables one to view more easily Monetarist vs. Keynesian arguments in dynamic terms; although, perhaps surprisingly, the crucial parameters turn out to be those emphasized in most static textbook models, R_{g^*}, R_m, and P'" (p. 746). These particular parameters are crucial in his model because they determine the speed of adjustment of income and thus the relative importance of the short run emphasized by the Keynesians and the long run emphasized by the Monetarists.

---

[3] A similar argument is presented less forcefully in Stein (1974, pp. 871–72).

[4] A somewhat similar argument can be based on the work of Edmund Phelps (chs. 2, 3, especially pp. 75–80).

[5] For the theoretical argument that workers are interested in after-tax wages see Thomas Dernburg, Robert Gordon (1971), and J. Johnston. For empirical confirmation see C. J. Bruce, Gordon (1971), and Johnston and M. Timbrell.

Stein, on the other hand, strongly emphasizes the sign of P_5, basically a long-run issue,[6] as *the* issue between the Keynesians and the Monetarists. (See his statement quoted above.) After responding to criticism of this position he reaffirmed it: "I am even more convinced now than I was that the sign of P_5...is the essence of the difference between the monetarists and the neo-Keynesians" (1976c, p. 258).

The question of exactly which issues are part of the Keynesian-Monetarist debate is perhaps a semantic one that can safely be left for future historians of economic doctrine. It does seem useful, however, to set out the following issues which are all part of the current debate over the mechanisms and effectiveness of macro-economic policies.

Short-run issues: Is it the case, as Van Order says, that "In the short run, emphasized by Keynesians, employment is variable and can be affected in the usual ways by monetary and fiscal policy" (p. 745)? Or is it the case that systematic monetary policy cannot affect output even in the short run, as argued by Thomas Sargent and Neil Wallace based on rational expectations (and variously rebutted by Phelps and John Taylor, Gordon (1977), and Stanley Fischer)? Or is the short run characterized by some other scheme whereby fiscal and monetary policy can affect output and unemployment, but in "unusual" ways?

Medium-run issues: Here, as Van Order emphasizes, the question is the magnitude of the speed of adjustment, which determines the relative importance of the short run and the long run.

Long-run issues: 1) Does bond-financed fiscal policy have any long-run effect on the growth of aggregate demand? This depends on P_5 in Stein's model and on R_θ in Van Order's. (See also Alan Blinder and Solow, 1973, 1976; Ettore Infante and Stein; James Tobin and Willem Buiter.) 2) Is fiscal or monetary policy capable, for example, for any of the three reasons suggested in Section III, of affecting the level of output and employment in the long run?

All of these questions will have to be satisfactorily resolved before macro-economic policy formation can be put on a truly firm footing.

---

[6] However, the sign of P_5 has medium-run implications (1976b, pp. 204–06).

REFERENCES

A. S. Blinder and R. M. Solow, "Does Fiscal Policy Matter?," *J. Publ. Econ.*, Nov. 1973, *2*, 319–37.

______ **and** ______, "Does Fiscal Policy Still Matter? A Reply," *J. Monet. Econ.*, Nov. 1976, *2*, 501–10.

C. J. Bruce, "The Wage-Tax Spiral: Canada 1953–1970," *Econ. J.*, June 1975, *85*, 372–76.

T. F. Dernburg, "The Macroeconomic Implications of Wage Retaliation Against Higher Taxation," *Int. Monet. Fund Staff Papers*, Nov. 1974, *21*, 758–88.

S. Fischer, "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule," *J. Polit. Econ.*, Feb. 1977, *85*, 191–205.

R. J. Gordon, "Inflation in Recession and Recovery," *Brookings Papers*, Washington 1971, *2*, 105–58.

______ "The Theory of Domestic Inflation," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 128–34.

E. F. Infante and J. L. Stein, "Does Fiscal Policy Matter?," *J. Monet. Econ.*, Nov. 1976, *2*, 473–500.

J. Johnston, "A Macro-Model of Inflation," *Econ. J.*, June 1975, *85*, 288–308.

______ **and M. Timbrell**, "Empirical Tests of a

Bargaining Theory of Wage Rate Determination," in David Laidler and David Purdy, eds., *Inflation and Labour Markets*, Manchester 1974.

Edmund Phelps, *Inflation Policy and Unemployment Theory*, New York 1972.

______**and J. B. Taylor,** "Stabilizing Powers of Monetary Policy under Rational Expectations," *J. Polit. Econ.*, Feb. 1977, *85*, 163–90.

T. Sargent and N. Wallace, "'Rational' Expectations, the Optimal Monetary Instrument and the Optimal Money Supply Rule," *J. Polit. Econ.*, Apr. 1975, *83*, 241–54.

R. M. Solow, "Comments on Stein," in Jerome Stein ed., *Monetarism*, Amsterdam 1976, 248–50.

Jerome Stein, "Unemployment, Inflation, and Monetarism," *Amer. Econ. Rev.*, Dec. 1974, *64*, 867–87.

______, (1976a) *Monetarism*, Amsterdam 1976.

______, (1976b) "Inside the Monetarist Black Box," in his *Monetarism*, Amsterdam 1976, 183–232.

______, (1976c) "Reply: A Keynesian Can Be a Monetarist," in his *Monetarism*, Amsterdam 1976, 253–71.

J. Tobin and W. Buiter, "Long-Run Effects of Fiscal and Monetary Policy on Aggregate Demand," in Jerome Stein, ed., *Monetarism*, Amsterdam 1976, 273–309.

R. Van Order, "Unemployment, Inflation, and Monetarism: A Further Analysis," *Amer. Econ. Rev.*, Sept. 1977, *67*, 741–46.

Unemployment, Inflation, and Monetarism—A Further Analysis: Reply

By ROBERT VAN ORDER*

I have no general comments on J. Kirker Stephens' insightful paper, with which I more or less agree. I do have five specific points.

1. Equation (8) in my paper is indeed incorrect, the constant term $R_m mn$ having been left out. The equations that follow ((9) and (10)) are correct, since the variables in them were defined to be measured as deviations from their long-run levels, which eliminates the constant terms.

2. The shift in the $Dg^*=0$ line (see my paper, p. 744) is a misprint and as suggested by Stephens should be a shift in the $Dy=0$ line.

3. I do not agree with the interpretation of my analysis of a change in fiscal policy (f). In my paper (as opposed to Jerome Stein's) the "goods" market clears instantaneously, and output can change instantaneously. In terms of the phase diagram, the initial level of y changes instantaneously when f changes. Hence, I do not argue (as in Stephens' fn. 2) that the rise in f shifts the $Dy=0$ line. Rather it raises the initial level of y. Subsequent adjustment of expectations moves the economy (f constant) into the northeast quadrant. Then when f is (instantaneously) reduced to its initial level we are left not at the "origin," but rather at a point like "D" in the diagram. I do think that this is an "interesting" experiment and Stephens' discussion, which emphasizes gradually adjusting f, says essentially the same thing. As is the case in all models that have a Phillips curve like equation (1) and have $g=g^*$ in the long run, neither fiscal nor monetary policy affect y in the long run unless they directly enter the Phillips curve.

4. With respect to long-run properties, Stephens is certainly correct that the $Dg^*=0$ line might not be horizontal, the role of tax rates and labor supply certainly being plausible. I still doubt that this is of central importance to the "Keynesian vs. Monetarist" controversy, because I doubt there is much disagreement about long-run properties of these models and because the dynamic properties appear to be reflected more in the $Dy=0$ line. In any event I doubt that the effects of, say, changing marginal tax rates on labor supply are qualitatively a major source of controversy.

5. Finally, as Stephens suggests, both my model and Stein's deal with the "old-fashioned" Monetarist controversy. Because they are built on an adaptive expectations approach to the adjustment of inflationary expectations they cannot, except trivially, deal with rational expectations. The only light we do have to shed is that if we view rational expectations as the limiting case where the speed of adjustment, λ goes to infinity, then such models are unstable.

REFERENCES

J. L. Stein, "Unemployment, Inflation and Monetarism," *Amer. Econ. Rev.*, Dec. 1974, *64*, 867–87.

J. K. Stephens, "Unemployment, Inflation, and Monetarism—A Further Analysis: Comment," *Amer. Econ. Rev.*, Sept. 1980, *70*, 732–37.

R. Van Order, "Unemployment, Inflation and Monetarism: A Further Analysis," *Amer. Econ. Rev.*, Sept. 1976, *67*, 741–46.

---

*U.S. Department of Housing and Urban Development.

On the "*q*" Theory of Investment

By HIROSHI YOSHIKAWA*

The purpose of this paper is to discuss the micro-economic foundations of the "*q*" theory of investment advanced by James Tobin, and by Tobin and William Brainard. This theory of investment can be summarized as follows:[1]

> The neoclassical theory of corporate investment is based on the assumption that the management seeks to maximize the present net worth of the company, the market value of the outstanding common shares. An investment project should be undertaken if and only if it increased the value of the shares. The securities markets appraise the project, its expected contributions to the future earnings of the company and its risks. If the value of the project as appraised by investors exceeds the cost, then the company's shares will appreciate to the benefit of existing stockholders. That is, the market will value the project more than the cash used to pay for it. If new debt or equity securities are issued to raise the cash, the prospectus leads to an increase of share prices.
> [Tobin and Brainard, p. 242]

> [Thus] the rate of investment — the speed at which investors wish to increase the capital stock — should be related, if to anything — to *q*, the value of capital relative to its replacement cost. [Tobin, p. 330]

> Economic logic indicates that a normal equilibrium value for *q* is 1 for reproducible assets which are in fact being reproduced, and less than 1 for others. Values of *q* above 1 should stimulate investment, in excess of requirements for replacement and normal growth, and values of *q* below 1 discourage investment.
> [Tobin and Brainard, p. 238]

The *q* theory of investment is now quite popular.[2] However, there seems to be considerable confusion about how the theory is to be interpreted. For example, Robert Hall argued that the *q* theory is basically a neoclassical theory à la Dale Jorgenson, and that only incomplete information and delivery lags can account for "disequilibrium" values of *q* and for their relation to investment. Otherwise, he argued, investment would keep *q* equal to one. On the other hand; Michael Lovell, "struck by similarities between *q* theory and the neoclassical approach of Dale Jorgenson, Robert Hall, and others (p. 399), criticized both approaches in the following way:

> Neither approach works out the dynamics of the adjustment process within the context of a carefully articulated optimization framework that would specifically incorporate the process of expectation formation and adjustment costs. In this respect, both theories are dominated by a number of contributions that have derived the optimal time path of adjustment simultaneously with the determination of the properties of the ultimate long-run equilibrium. [pp. 399–400]

The major point of this paper is that *q* theory can be derived from a choice-theoretic framework which explicitly takes account of adjustment costs associated with investment. The framework is similar to those of Robert Lucas, John Gould, and

---

*State University of New York-Albany. I would like to thank James Tobin, Katsuhito Iwai, Hirofumi Uzawa, and an anonymous referee for their helpful comments.

[1] Elaborations and qualifications of the theory can be found in section 1 of Tobin and Brainard.

[2] For example, Robert Hall called it "the major competitor to Jorgenson's theoretical framework for investment" (p. 85).

Hirofumi Uzawa, which were in turn anticipated by Robert Eisner and Robert Strotz. These authors criticized Jorgensonian investment theory on the following grounds. In Jorgenson's theory, "desired" capital stock is first determined and then investment is derived as a process induced by the discrepancy between the desired capital stock and current actual capital stock. The theory assumes, of course, the existence of adjustment costs associated with investment. Otherwise, Jorgenson's theory can explain only optimal capital stock, but not optimal investment. However, if the existence of adjustment costs is important, it must be taken into account explicitly in firms' decisions about investment. In this case, the notion of desired capital stock derived apart from adjustment costs loses much of its meaning.

In what follows, I show that the q theory of investment can be derived following Lucas-Gould-Uzawa models of investment which emphasize the role of adjustment costs. For this purpose, let us consider a one-commodity model in which the money price of this single commodity is denoted by p. Suppose that the production technique of the firm under consideration is described by the following production function:

$$Q=F(L,K) \tag{1}$$

where Q, L, and K are, respectively, output, labor employed, and an index of fixed capital such as machines, factories, and managerial resources. The function F is assumed to be homogeneous of degree one. Given K, the firms' labor employment decision can be characterized in the usual way:

$$\underset{L}{\mathrm{Max}} \int_0^\infty (pQ-wL)e^{-\rho t}\,dt \tag{2}$$

where w is the wage rate and ρ is the rate of return on capital equity required by stockholders. Given expected p and w, the necessary condition is[3]

$$p(\partial Q/\partial L)=w \qquad \text{for all t} \tag{3}$$

Under the homogenity assumption of F, this is equivalent to

$$f(k)-f'(k)k=w/p \tag{4}$$

where $f(k)=F(K/L,1)$ and $k=K/L$. Denoting the expected marginal product of capital, or the expected rate of profit as r, we have

$$r=f'(k) \tag{5}$$

Thus, under the optimal employment decision (4) corresponding to a given expected w and p, the stock price of the firm (2) is reduced to the following:

$$\int_0^\infty prKe^{-\rho t}dt=prK/\rho \tag{6}$$

Let us consider a current "once and for all" investment decision of the firm. The firm must decide how fast it should increase its capital stock. An increase in capital is denoted by ΔK. As for a decrease in capital, there is usually a lower limit set by natural depreciation, since physical capital is not malleable and the firm incurs prohibitive costs when decreasing the existing capital stock too fast. (Kenneth Arrow analyzes the case where investment is irreversible in an explicitly dynamic model.) I will assume $\Delta K \geqslant 0$ in what follows, and for simplicity, depreciation is ignored.

Now the fundamental factor required to eliminate the indeterminancy of the optimal level of investment is the adjustment cost associated with high investment, that is, with a rapid rate of capital increase. It should be clear that this point is related to the distinction between the marginal efficiency of capital and the marginal efficiency of investment. Let us assume that the effective price (after taking into account adjustment costs) of investment goods is not lower than the going market price of capital goods p, and that it increases as the amount of investment increases. This is due to the existence of a particular kind of resource such as management which cannot be increased proportionally to physical capital and/or to an increasing supply price of investment goods. Specifically, I will introduce the following function:

---

[3] For simplicity, I assume static expectations about p and w. The basic analysis, however, remains unchanged even if I assume changing expected p and w in so far as they are expected to change at the common trend rate of inflation.

$$\psi(\alpha) \tag{7}$$

where $\psi(0)=1, \psi'>0, \psi''>0, \alpha=\Delta K/K$

$$\lim_{\alpha\to 0+} \psi'(\alpha)=0$$

Under this assumption, the effective price of investment ΔK when the growth rate of capital is α, is not $p\Delta K$ but $p\psi(\alpha)\Delta K$, which is larger than $p\Delta K$. Then the market value of this firm after the investment is the following:

$$\begin{aligned} (8)\quad & pr(K+\Delta K)/\rho-p\psi(\alpha)\Delta K \\ & \quad =pr(1+\alpha)K/\rho-p\psi(\alpha)\alpha K \end{aligned}$$

The firm is assumed to maximize this value with respect to $\alpha \geqslant 0$. The first-order condition is

$$r/\rho-\psi'\alpha-\psi \leqslant 0 \tag{9}$$

where, if $\alpha>0$, the equality holds. The second-order condition $-2\psi'-\alpha\psi''<0$ is always satisfied. From this, we know that for $\Delta K>0$, $q\equiv r/\rho>1$ is necessary. Note that in the present framework the value of capital stock evaluated in the financial market prK/ρ, relative to its replacement cost pK, is r/ρ. This is Tobin's q. If $r/\rho \leqq 1, \Delta K=0$ in this model. If we now define $f(\alpha)=\psi'(\alpha)\alpha+\psi(\alpha)$ where $f'>0$ and $f(0)=1$, it can be seen from Figure 1 that the optimal growth rate of capital stock α^*, and thereby the optimal level of investment $\Delta K=\alpha^*K$, are uniquely determined. Note that $f(\alpha)$ in Figure 1 indicates the *marginal* effective cost of capital increase expressed as a function of the relative rate of capital increase, and that the condition for optimal α^* is the equality of q to this marginal effective cost, $f(\alpha)$. Thus given the marginal effective cost function $f(\alpha)$, investment is shown to be an increasing function of q in Figure 1.

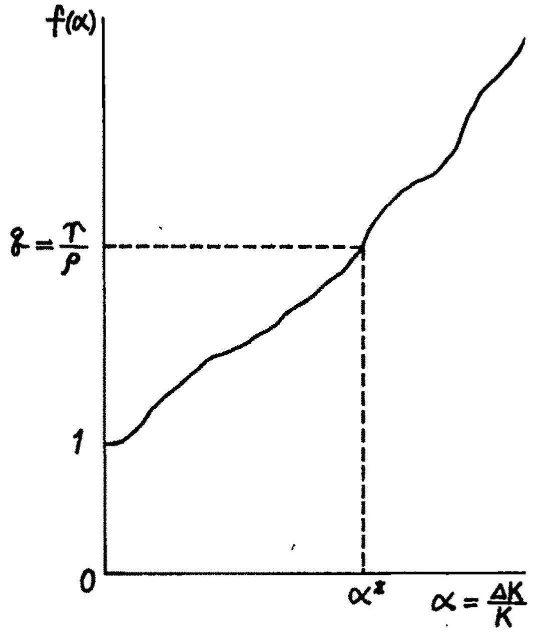


FIGURE 1

Before proceeding to a more complicated case, some qualifications about the above analysis are in order. One is the specific nature of the adjustment cost function $\psi(\alpha)$ (equation (7)). Although a convex adjustment cost function is fairly standard in the literature (for example, see Lucas, Gould, and Uzawa), Michael Rothschild emphasized the possibility of a concave adjustment function. More general functions might affect the above analysis. Secondly, the fact that r is taken as given is valid only under our simplifying assumption of a homogeneous production function. Under decreasing returns to scale r decreases when K increases. This would not, however, affect the above analysis substantially. Finally, the expected return on new capital is assumed to be equal to that on existing capital r. If they are not equal, it is "the q ratio *on the margin* that matters for investment: the ratio of the increment of market valuation to the cost of the associated investment" (Tobin and Brainard, p. 243). This can be seen in my analysis by assuming the following instead of (8):

$$(10)\quad prK/\rho+pr'\Delta K/\rho-p\psi(\alpha)\Delta K$$
$$=prK/\rho+pr'\alpha K/\rho-p\psi(\alpha)\alpha K$$

This is the market value of the firm when the expected return on new capital, r' is different from r. The maximization of (10) with respect to α leads to the optimality condition which requires the equality of $q' = r'/\rho$ to $f(\alpha)$. That is, the *marginal* q, q' must be equal to the *marginal* effective cost of capital increase, $f(\alpha)$. The reinterpretation of Figure 1 should be straightforward. These qualifications are also applied to a dynamic model to which I now turn.

The above static model should make clear my basic point but it assumed, among other things, that the investment decision considered is once and for all. More generally, however, investment decisions would be made successively, and in this case they must be made optimally from a dynamical viewpoint. Mathematically the problem is one of calculus of variations rather than differentiation.[4] Uzawa analyzes this more complicated problem. The simplified version of his model is

$$(11)\quad \underset{z(t)}{\text{Max}}\int_0^\infty p(r-\Phi(z))K(t)e^{-\rho t}dt$$

where $z(t)=DK(t)/K(t)$ and r, ρ, $K(0)$, p are all given. The term $\Phi(z)$ is the effective cost of adding capital after taking account of adjustment cost. This effective cost is greater the faster is the growth rate of capital. Specifically we assume $\Phi(z)$ has the following property:

(12)

$$\Phi'(z)>0, \Phi''(z)>0, \Phi(0)=0, \Phi'(0)=1$$

Uzawa's analysis shows that if a solution exists, then it is unique and the optimal z is constant over time. It is also shown that the optimal z, z^*, must satisfy the following relation:

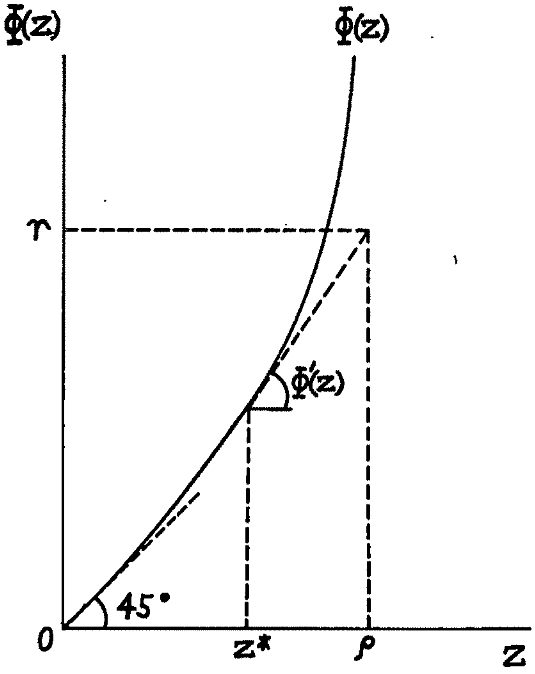


FIGURE 2

$$(13)\quad (r-\Phi(z^*))/(\rho-z^*)=\Phi'(z^*)$$

Thus z^* can be found from Figure 2. Note that $(r-\Phi(z))/(\rho-z)$ is q in this dynamic context since on the optimum path z is constant and so from (11) the present value of the firm evaluated in the financial market is $p(r-\Phi(z))K(0)/(\rho-z)$, whereas its replacement cost is $pK(0)$. Therefore q is modified by the growth rate of the firm, z, and the steady investment payment $\Phi(z)$ in this dynamic model. The slope of $\Phi(z)$ is, on the other hand, the *marginal* effective cost of investment. It should be clear from Figure 2 that only at z^* is the optimality condition satisfied: $q=(r-\Phi(z))/(\rho-z)$ is equal to the *marginal* effective cost of investment, $\Phi'(z)$. This is a natural extension of the previous static result. It should also be clear that given the adjustment cost function $\Phi(z)$, and present capital stock K, investment (or z) is an increasing function of q.

One point to be noted here is the meaning of the assumption $\Phi'(0)=1$. This assumption means that the effective cost of adding one additional unit of capital is simply the market price of capital goods. If there exists some fixed cost associated with any investment, then we would have $\Phi'(0) > 1$ instead of $\Phi'(0) = 1$, and in this case the critical level of q would be larger than one.

---

[4] This is the problem analyzed by Lucas, Gould, and Uzawa.

Let me summarize my analysis. In a simple investment model emphasizing adjustment cost (this seems to be a reasonable one if we recognize that investment decisions concern *the speed of increase* in capital stock), it has been shown that the optimality condition requires the equality of q (the *marginal* q if the expected return on new capital is different from that on existing capital) to the *marginal* effective cost of investment. Aside from the price of capital goods, the latter is determined by the adjustment cost function. Therefore, the above optimality condition means that given the shape of the adjustment cost function, investment is an increasing function of q. Thus, contrary to Lovell's comment on the q theory of investment, this theory can be derived from the maximizing behavior of the firm explicitly taking into account adjustment cost. By implication, this paper has also shown that the q theory is different from Jorgenson's neoclassical theory. The q theory, allowing the divergence between the value of capital evaluated in the financial market and the price of capital goods, is a theory which explains how investment (change in capital stock) is motivated by this apparent short-run disequilibrium. No wonder, as observed above, adjustment cost plays a crucial role in this theory. Jorgenson's theory, on the other hand, basically concerns the long-run demand for capital stock.

One final remark is that in the q theory, as in John Maynard Keynes' exposition in chapter 12 of his *General Theory*, the rate of profit r is the one expected by stockholders. It seems worth future study to explore whether or not the divergence between expectations of r held by asset holders and by entrepreneurs has some implications for investment theory.[5]

REFERENCES

K. J. Arrow, "Optimal Capital Policy with Irreversible Investment," in J. N. Wolfe, ed., *Value Capital and Growth: Papers in Honour of Sir John Hicks*, Edinburgh 1968.

R. Eisner and R. H. Strotz, "Determinants of Business Investment," in *Impacts of Monetary Policy: Commission on Money and Credit*, Englewood Cliffs 1963, 59–337.

J. P. Gould, "Adjustment Costs in the Theory of Investment of the Firm," *Rev. Econ. Stud.*, Jan. 1968, *35*, 47–55.

R. E. Hall, "Investment, Interest Rates, and the Effects of Stabilization Policies," *Brookings Papers*, Washington 1977, *1*, 61–103.

D. W. Jorgenson, "Anticipations and Investment Behavior," in James S. Duesenberry et al., eds., *The Brookings Econometric Model of the United States*, Amsterdam 1965, 35–92.

John Maynard Keynes, *The General Theory of Employment, Interest and Money*, London 1936.

M. C. Lovell, "Comments and Discussion of von Furstenberg's Paper," *Brookings Papers*, Washington 1977, *2*, 398–401.

R. E. Lucas, Jr., "Adjustment Costs and the Theory of Supply," *J. Polit. Econ.*, Aug. 1967, *75*, 321–34.

M. Rothschild, "On the Cost of Adjustment," *Quart. J. Econ.*, Nov. 1971, *75*, 605–22.

J. Tobin, "A General Equilibrium Approach to Monetary Theory," in his *Essays in Economics: Macroeconomics*, Vol. 1, Chicago 1971, 322–38.

______ **and W. C. Brainard,** "Asset Market and the Cost of Capital," in Bela Balassa and Richard Nelson, eds., *Economic Progress, Private Values and Public Policy: Essays in Honor of William Fellner*, Amsterdam 1977, 235–62.

H. Uzawa, "Time Preference and the Penrose Effect in a Two-Class Model of Economic Growth," *J. Polit. Econ.*, July/Aug. 1969, *77*, Part II, 628–52.

---

[5] One possible answer would be that if the two expectations are different, firms buy and sell their own stocks, so that the divergence between two expectations has not much significance. Cambridge economists, on the other hand, seem to emphasize the role of expectations of entrepreneurs distinguished from those of asset holders.

Optimal Corrective Taxes or Subsidies when Revenue Raising Imposes an Excess Burden

By YEW-KWANG NG*

This paper addresses the following problem. Suppose that for some reason, a tax or subsidy on a certain good may be desirable, for example, as it produces an external diseconomy or economy. However, a subsidy has to be financed by other taxes and a tax contributes to government revenue. If the revenue can be raised by nondistortive taxes, then this may not be a relevant issue. However, revenue has to be raised by distortive taxes in practice (usually income or value-added taxes). Is it then still desirable to tax or subsidize external diseconomies or economies?[1] One is inclined to reply that it all depends on the relevent interrelationships and nothing can be said a priori due to the second best considerations. In this paper, it is argued that, given a reasonable assumption likely to prevail in most cases, a normal tax or subsidy is still desirable provided the government revenue requirement is not too high. The "counternormal" policy of taxing/subsidizing external economies/diseconomies cannot however be ruled out completely, even if the amount of revenue required is zero.

This problem has been largely ignored in the literature. The optimal taxation literature (for an introduction, see Agnar Sandmo, 1976, or Anthony Atkinson, 1977; for an authoritative synthesis, see James Mirrlees, 1976) concentrates on taxation for revenue and income distribution purposes without incorporating corrective taxes or subsidies. On the other hand, the literature on the Pareto optimality of a competitive equilibrium (with taxes or subsidies) in the presence of externalities usually ignores the government revenue requirement and the impracticability of nondistortive taxes. (See, for example, Hiroaki Osana where lump sum taxes and subsidies are used.)[2]

In the following analysis, I assume that the government revenue requirement is given and has to be raised by a proportional income taxation. Income taxation is selected partly for its practical importance and partly for the fact that, if we abstract from differential rates (which may be explained by such considerations as merit/demerit goods, externalities, etc. themselves), general commodity taxation (for example, a value-added tax) is equivalent to a proportional income tax. My disregard for possible progressivity is due mainly to the need for analytical simplicity, but may be justified by a predominant concern with efficiency issues.[3] While I keep referring to externalities, my analysis may also be applied to other factors justifying corrective taxes and subsidies. For

---

*Monash University. I am grateful to the managing editor and a referee for helpful suggestions.

[1] It is true that, for a given situation, we do not have to subsidize an external economy, we could tax it negatively; i.e., the less the activity, the higher the total tax bill. This is equivalent to a subsidy plus a lump sum tax. The government does not then have to finance the subsidy. However, while this may be good enough for this particular externality, it will adversely affect efficiency in the long run as people will then refrain from producing external economies. Moreover, the point of reference to determine the negative tax or the amount of the lump sum tax is difficult to determine in an acceptable way. The negative tax solution to the problem may be safely dismissed as both inefficient and impractical.

[2] See, however, Efraim Sadka who derives optimality conditions for taxes with consumption externalities. Consistent with the theory of second best, the direction of taxes and subsidies can go either way. Sandmo (1975) also analyzes a similar problem in a different (commodity taxation) framework than mine. My income taxation framework recognizes the practical difficulties of a system of complicated rates of commodity taxes and leads to the concentration on factors such as the degress of substitutability with leisure. I am indebted to Richard Zeckhauser for drawing my attention to the last reference.

[3] Linearity in taxes but no strict proportionality also receives support from the optimal income taxation literature (see Mirrlees, 1971, and Atkinson, 1973). My analysis can be revised to cover the linear case.

example, the analysis below can be interpreted, without any substantive changes, to refer to merit/demerit goods.

For the benefit of nonspecialist readers, the main results of this paper are presented in Section I. The analysis leading to these results is contained in Section II. The last section remarks on the policy relevance of my analysis and the relaxation of some of the simplifying assumptions used.

I. Results

Using the usual simple model of the optimal commodity taxation literature and reasonably assuming that the social marginal utility of income is positive, the following proposition may be derived.

PROPOSITION 1: *In the simple economic model where the government revenue must be raised by distortive income and/or uniform value-added taxes, the presence of an external diseconomy/economy still calls for a corrective tax/subsidy provided that* (a) *an increase in the (consumer) price of the externality-producing good is more effective in reducing its consumption proportionately than is an increase in the (consumer) price of labor in increasing it, proportionately to labor*; (b) *a compensated increase in the price of labor is more effective in increasing labor supply proportionately than is a compensated increase in the price of the externality-producing good in reducing labor supply, proportionately to the amount of the good;*[4] (c) *the government revenue requirement is not too high. If the revenue requirement is not negligible, the case for a subsidy is weakened and may be reversed but the case for a tax is strengthened.*

After the theory of second best, it is not surprising that a corrective tax/subsidy based on the partial-equilibrium analysis may no longer be desirable in the presence of the need to raise distortive taxes. For example, consider a subsidy for an external economy which must be financed by an increase in income tax. If condition (a) in Proposition 1 is reversed, a reduction in the price of the good due to the subsidy is less effective in increasing its consumption than a reduction in the price of labor (due to an increase in income tax) in reducing its consumption, proportionately. The subsidy (and the required income tax to finance it) may then be counterefficient. However, as argued in Section II, both conditions (a) and (b) are likely to hold for most goods. Moreover, if either condition is violated, it is likely that the other one does not hold either. Normal corrective taxes/subsidies still apply due to the following proposition.

PROPOSITION 2: *Proposition* 1 *remains true if both conditions* (a) *and* (b) *are reversed.*

Even for cases where only one of the two conditions are violated, it may still be desirable to adopt normal corrective taxes/subsidies. But the counternormal policy of taxing/subsidizing goods producing external economies/diseconomies cannot be ruled out completely. Due to the fact that conditions (a) and (b) can be expressed in uncompensated forms, the empirical task of establishing their fulfillment is made less difficult.

II. Analysis

Let us first analyze a deliberately simplified model so that the central argument can be appreciated easily. Suppose that there are $n+1$ commodities in the economy, the last one being labor. As in the optimal commodity taxation literature, assume that the consumer side of the economy can be represented by one consumer or a community utility function,

$$U(x^1, x^2, \dots, x^n, L, E) \tag{1}$$

where x^i is the amount of the ith good consumed, L is labor, and E is the amount of the only external effect produced either by the consumption or production of the first good. Abstracting from the problem of increasing or reducing E for any given x^1, we may define the units such that $E=x^1$. Each consumer is taken as atomistic and takes prices, tax rates, wage rate, and E as given (not affected by his own decision). He maximizes (1) with respect to L and x^i subject to

---

[4] Observant readers may have noticed that condition (a) is stated in the uncompensated form while condition (b) is in the compensated form. Due to the precise proportionalities involved, the income effects cancel out exactly. Hence if either condition holds in the compensated form, it also holds in the uncompensated form and vice versa (see Section II). I intentionally state (b) in the compensated form, as the uncompensated form may appear less plausible than it really is.

$$(2)\quad (1+t)p^1x^1+\sum_{2}^{n}p^ix^i=(1-T)wL$$

where t is the tax rate (subsidy if negative) on the first good, T is the tax rate on income, p^i are prices, and w is the wage rate. Thus, only a proportional income tax (or a uniform value-added tax) and a possible tax/subsidy on good 1 are being considered. The first-order conditions for a maximum are

$$(3a)\qquad U_1=\lambda(1+t)p^1$$

$$(3b)\qquad U_i=\lambda p^i \qquad i=2,\dots,n$$

$$(3c)\qquad -U_L=\lambda(1-T)w$$

where λ, being the Lagrangean multiplier associated with (2), is the marginal utility of income. A subscript denotes partial differentiation, for example, $U_i\equiv\partial U/\partial x^i$. The term U_1 takes into account only the internal effect of x^1 upon U and does not include the external effect through E.

Again as in the basic optimal tax model, producers' prices and the wage rate are taken to be fixed. Such might occur in a small trading nation producing nontraded goods with constant returns. The society is imagined to maximize (1) with respect to t and T, subject to a fixed-revenue requirement

$$(4)\qquad TwL+tp^1x^1=R$$

The following first-order conditions may be derived:

$$(5)\quad \sum_{i=1}^{n}U_ix_t^i+U_LL_t+U_Ex_t^1=\theta\left(p^1x^1+wTL_t+tp^1x_t^1\right)$$

$$(6)\quad \sum_{i=1}^{n}U_ix_T^i+U_LL_T+U_Ex_T^1=\theta\left(wL+wTL_T+tp^1x_T^1\right)$$

where θ is the Lagrangean multiplier associated with (4).

Differentiate (2) with respect to t and T to obtain, respectively,

$$(7)\quad (1+t)p^1x_t^1+\sum_{i=2}^{n}p^ix_t^i-(1-T)wL_t=-p^1x^1$$

$$(8)\quad (1+t)p^1x_T^1+\sum_{i=2}^{n}p^ix_T^i-(1-T)wL_T=-wL$$

Substitute U_i and U_L from (3) into (5) and (6), and then substitute (7) and (8), respectively, into the resulting equations, yielding

$$(9)\quad -\lambda p^1x^1+U_Ex_t^1=\theta\left(p^1x^1+wTL_t+tp^1x_t^1\right)$$

$$(10)\quad -\lambda wL+U_Ex_T^1=\theta\left(wL+wTL_T+tp^1x_T^1\right)$$

Now in the consumer budget constraint (2), $(1+t)p^1\equiv q^1$ is the consumer price of x^1 and $(1-T)w\equiv q^L$ is the consumer price of labor/leisure. Thus we may write

$$(11)\quad x_t^1=\frac{\partial x^1}{\partial q^1}\cdot\frac{\partial q^1}{\partial t}=p^1\left(S_{11}-x^1\frac{\partial x^1}{\partial I}\right)$$

$$(12)\quad x_T^1=\frac{\partial x^1}{\partial q^L}\cdot\frac{\partial q^L}{\partial T}=-w\left(S_{1L}+L\frac{\partial x^1}{\partial I}\right)$$

where $S_{11}\equiv\partial x^1/\partial q^1|_U$ is the substitution effect on x^1 of an increase in q^1, $S_{1L}\equiv\partial x^1/\partial q^L|_U$, and $I=$income. Similarly, we have

$$(13)\quad L_t=\frac{\partial L}{\partial q^1}\cdot\frac{\partial q^1}{\partial t}=p^1\left(S_{L1}-x^1\frac{\partial L}{\partial I}\right)$$

$$(14)\quad L_T=\frac{\partial L}{\partial q^L}\cdot\frac{\partial q^L}{\partial T}=-w\left(S_{LL}+L\frac{\partial L}{\partial I}\right)$$

Substitute (11)–(14) into (9) and (10), and write in proportional form to eliminate the

Lagrangean multiplier θ, giving

$$(15)\quad \frac{p^1x^1A-p^1S_{11}U_E/\lambda}{wLA+wS_{1L}U_E/\lambda} = \frac{p^1x^1B+p^1(wTS_{L1}+tp^1S_{11})}{wLB-w(wTS_{LL}+tp^1S_{1L})}$$

where $A\equiv 1+(U_E/\lambda)\cdot(\partial x^1/\partial I)$ and $B\equiv 1-wT(\partial L/\partial I)-tp^1(\partial x^1/\partial I)$. Assuming that the second-order condition is satisfied, (15) and (4) determine the optimal values of t and T.

It can be shown that both A and B can be taken as positive. Being the private (i.e., not accounting for externality) marginal utility of purchasing power, λ can be taken as positive. The term U_E is positive/negative if the externality is an economy/diseconomy. Thus, if x^1 is not an inferior good, A is necessarily positive for the case of an external economy. Similarly, if x^1 is an inferior good, A is necessarily positive for the case of an external diseconomy. Ambiguities arise only for a noninferior good with external diseconomy and for an inferior good with external economy. However, A is positive (though less than one) in both cases unless the externality is so strong that an increase in income actually makes the community worse off. It is reasonable to assume that $-U_E(\partial x^1/\partial I)<\lambda$. In other words, while an increase in income may reduce utility through the externality effect, the reduction does not overbalance the internal effect λ. If this is not true, an increase in income actually makes the community worse off. It is then not difficult to see that these conclusions may then have to stand on their heads, since an improvement in economic efficiency may then make things worse. It is thus clear that $A>0$ is a reasonable assumption.

Turning to examine $B(\equiv 1-wT(\partial L/\partial I)-tp^1(\partial x^1/\partial I))$, it is clearly positive when $T=t=0$. Even if the tax/subsidy rates are nonzero, a positive B just means that a dollar increase in income does not induce (through income effects) an increase in tax revenue by a dollar or more. In fact, as leisure can be taken not as an inferior good, $-\partial L/\partial I\geqslant 0$. For B to be negative, $tp^1(\partial x^1/\partial I)$ would have to be larger than one which is quite impossible, unless t is very large (or, if $\partial x^1/\partial I<0$, t is negative and large in absolute value). Since I am going to argue just for a positive/negative t (for the case of external diseconomy/economy), the case where t is already very large (absolutely) to begin with is not relevant to the argument. It is thus perfectly legitimate to take $B>0$.

To abstract from the complication of a positive revenue requirement, let us for the moment take $R=0$. This does not assume the problem away completely since for a change in t (negative t in particular), T has to be changed to meet the requirement $R=0$. It is thus not obvious that $t\gtrless 0$ is optimal if $U_E\lesseqgtr 0$. In fact, it can be seen that there are cases (though unlikely to prevail) where the sign of optimal t is reversed (i.e., becomes the same as U_E) and there are borderline cases where $t=T=0$ is optimal even if $U_E\neq 0$. For this last one to be true, it can be observed from (15) that $-p^1S_{11}/wS_{1L}=p^1x^1/wL$ must hold,[5] or

$$(16)\qquad -S_{11}/x^1=S_{1L}/L$$

Assuming that the usual postulates of consumer theory are satisfied, S_{11} is necessarily negative. Thus (16) can be satisfied only if $S_{1L}>0$. But from the general theory of consumption, S_{1L} can be of any sign. Thus, at a general level, it is possible for (16) to hold and for $t=T=0$ to be optimal even if $U_E\neq 0$. In fact, if $-S_{11}/x^1<S_{1L}/L$, it *may* be optimal to pursue the counternormal policy of subsidizing an external diseconomy and taxing an economy. The reason for this is not difficult to see. For efficiency purposes, it is the substitution effect that counts. Take the case of an external economy. A subsidy on x^1 (which reduces q^1) must be financed by an increase in income tax T (which reduces q^L). But if, as

---

[5] When $t=T=0$, the right-hand side of (15) reduces to p^1x^1/wL. On the left-hand side, since A is associated with p^1x^1 on the numerator and with wL on the denominator, we may take out the terms associated with A without affecting the proportionality. Specifically, (15) simplifies into (16) after putting $t=T=0$ and using cross multiplication.

far as substitution effects are concerned, the reduction in q^L is more effective in reducing x^1 than is the reduction in q^1 in increasing x^1 proportionately, the subsidy may be positively harmful. Hence, a tax on x^1 enabling a lowering in T may be desirable instead. However, it is reasonable to assume that, in normal cases,

$$(17) \qquad -S_{11}/x^1 > S_{1L}/L$$

which says that a compensated increase in q^1 is more effective in reducing x^1 proportionately than a compensated increase in q^L in increasing (if at all) x^1, proportionately to L. This is likely to hold as a change in the consumer price of the first good itself can be more effective proportionately, in affecting the consumption of x^1, than a change in the price of something else. One might think that this may not be so since the something else which happens to be labor may be much more important than x^1. But x^1 and L appear in the denominators of (7), hence the effect due to the relative importance of the two goods has already been taken into account.

Inequality (17) may be expressed in a different way. Due to the symmetry of substitution effects, $S_{1L} = -S_{L1}$ (the negative sign accounts for the fact that an increase in leisure is a reduction in labor), (17) may be rewritten as

$$(17') \qquad -\frac{q^1}{x^1}\cdot\frac{\partial x^1}{\partial q^1}\Big|_U > -\frac{q^1}{L}\cdot\frac{\partial L}{\partial q^1}\Big|_U$$

which says that, as far as substitution effects are concerned, x^1 is more elastic than L with respect to a change in q^1 (the consumer price of x^1) itself. (Taking $\partial L/\partial q^1|_U$ to be negative; if it is positive, the inequality is necessarily satisfied.) A compensated increase in q^1 reduces x^1 and increases the consumption of the composite good $(x^1,\ldots,x^n+\text{leisure})$. Hence, $\partial L/\partial q^1|_U$ tends to be negative. But in proportionate terms, this secondary effect is unlikely to be more important than the primary effect on x^1 itself. The reduction in x^1 concentrates on x^1 itself while the increase in the composite good is spread across (not necessarily evenly and positively) all other goods and leisure. Hence, leisure is likely to increase by only a small amount in absolute terms and even less in proportionate terms if x^1 is less important than leisure in consumption.

It may be thought that (17′) may well be reversed since the left-hand side, though nonnegative, may be very small. In the limiting case, it approaches zero. In this case, is not the right-hand side quite likely to exceed the left-hand side? The answer is negative. As the reduction in x^1 approaches zero, the increase in the composite good approaches zero. The right-hand side is still likely to be smaller than the left-hand side due to the argument of the preceding paragraph. If we allow both the left-hand side and the right-hand side of (17′) to become exactly zero, then of course (17′) becomes an equality (i.e., equation (16)) and the optimality condition (15) is satisfied at $t=T=0$. This result is not surprising; with no possibility for substitution, a tax/subsidy serves no efficiency-improving function.

While (17′) can be shown to be true in most cases, its reversal cannot be ruled out completely. This may happen if the consumption of x^1 is not time intensive and the close substitutes of x^1 are very time intensive. Then a reduction in x^1 does not release much time, but the increase in the consumption of its substitutes takes up more time, leaving less time for working. Then $\partial L/\partial q^1|_U$ may be very negative, leading to a possible (but unlikely) reversal of (17′). The counternormal policy of taxing/subsidizing external economy/diseconomy *may* then be desirable. The requirement that x^1 be non-time intensive may appear counterintuitive. Take the case of an external economy. A subsidy increases x^1 which, if time intensive, may lead to a reduction in L and may thus be undesirable as revenue is dependent on L. This apparent paradox can be explained. A subsidy on x^1 has to be financed by an increase in the income tax rate which reduces the consumer price of time (labor/leisure). This tends to encourage the consumption of more time-intensive goods and less non-time-intensive goods. If x^1 is non-time intensive and its substitutes are time intensive, the consumption of x^1 may be reduced, destroying the purpose of the subsidy. However, unless there is some specific reason to believe that the time intensities and other possible relevant factors are very biased, it is reasonable to take (17) as satisfied. With this reasonable assumption (and a similar one to be mentioned below), I can show that some tax/subsidy on external diseconomy/economy is desirable.

If (17) holds, (16) cannot hold and $t=T=0$ does not satisfy the optimality condition (15). The left-hand side of (15) will be larger or smaller than p^1x^1/wL, according to whether U_E is positive or negative. Let us consider the case $U_E>0$. (The opposite case can be similarly handled.) In this case of external economy, (15) can be satisfied by $t<0$ (subsidy on x^1) and $T>0$ (to meet the revenue constraint). To see this, observe the right-hand side of (15). The term associated with t is $p^1(p^1S_{11})$ in the numerator and $-p^1(wS_{1L})$ in the denominator. Thus if (17) holds, a negative t makes the right-hand side larger than p^1x^1/wL and hence tends to make it equal to the left-hand side. The terms associated with T are $w(p^1S_{L1})$ in the numerator and $-w(wS_{LL})$ in the denominator. $S_{LL}\equiv\partial L/\partial q^L|_U$ is necessarily positive. Similar to (17), it is reasonable to assume

$$S_{L1}/x^1>-S_{LL}/L \tag{18}$$

which says that a compensated increase in q^L is more effective in increasing L proportionally than is a compensated increase in q^1 in reducing L in proportion to x^1. Or, due to the symmetry of substitution effects, $-S_{1L}/x^1>-S_{LL}/L$, which says that a compensated increase in q^L is more effective in increasing L than it is in increasing x^1 proportionately. Then an increase in T also makes the right-hand side of (15) larger than p^1x^1/wL. Thus starting from a position of $T=t=0$, making t negative and T positive tends to satisfy (15), since, in the present case of external economy, the left-hand side of (15) is larger than p^1x^1/wL. Similarly, in the case of diseconomy, the left-hand side of (15) is smaller than p^1x^1/wL. Then making t positive and T negative tends to achieve (15).

If the government revenue requirement is not negligible, T may be positive to begin with. Then the right-hand side of (15) is already larger than p^1x^1/wL even at $t=0$. The case for a subsidy on external economy cannot be established without a detailed comparison of the values of the relevant substitution effects, external effects, etc. The case for a tax on external diseconomy is made even stronger, as may be expected intuitively.

The discussion above may be summarized into the following proposition:

PROPOSITION 1′: *In the simple economic model where the government revenue must be raised by distortive income and/or uniform value-added taxes, the presence of an external diseconomy/economy still calls for a corrective tax/subsidy provided that* (a) *a compensated reduction in the consumer price of the externality-producing good is more effective in increasing the consumption of this good than it is in increasing the labor supply, proportionately*; (b) *a compensated increase in the consumer price of labor is more effective in increasing the labor supply than it is in increasing the consumption of the externality-producing good, proportionately*; (c) *the government revenue requirement is not too high. If the revenue requirement is not negligible, the case for a subsidy is weakened and may be reversed but the case for a tax is strengthened.*

In fact, Proposition 1′ has not gone far enough. (It is nevertheless preserved for its simplicity; it is sufficient for normal cases.) Suppose that condition (a) in Proposition 1′ (i.e., inequality (17) or (17′)) is reversed, does this mean that the counternormal policy of taxing/subsidizing external economy/diseconomy is then desirable? Not necessarily. (Observant readers may have noticed that I say the counternormal policy *may* be desirable.) Now the left-hand side of (15) is made smaller than p^1x^1/wL for the case of external economy. But a negative t (i.e., subsidy) also makes the right-hand side smaller. If condition (b) (i.e., inequality (18)) is also reversed, the required increase in T to finance the subsidy also makes the right-hand side smaller. It is then clear that the normal policy of taxing/subsidizing external diseconomy/economy is still desirable. The reversal of both the two normality conditions results in normality. This can be explained intuitively. A subsidy on x^1 (which reduces q^1 and increases x^1) must be financed by an increase in T which reduces q^L. If (17) is reversed, a reduction in q^L greatly reduces x^1 and may destroy the purpose of the subsidy. However, if (18) is also reversed, the reduction in q^1 increases L even more greatly, making only a small increase in T sufficient to raise enough revenue to finance the subsidy. Though an increase in T greatly reduces (through q^L) x^1, this effect may not be sufficient to negate the case for a subsidy as the required increase in T (if at all) is small.

The above discussion may be summarized into the following proposition:

PROPOSITION 2′: *Proposition* 1′ *remains true if both conditions* (a) *and* (b) *are reversed.*

For cases where only one of (a) and (b) is reversed, it is not necessarily true that the counternormal policy follows. For example, if (a) is strongly satisfied and (b) is just marginally reversed, normal policy may still be desirable. In these complicated (but unlikely) cases, we have to look into the relevant values in more detail. We cannot however completely rule out the counternormal cases even by resort to second-order conditions as when either q^1 or q^L changes, all goods including $x^2, \ldots, x^n$ may change, making the values of S_{1L}, S_{LL}, etc. quite flexible. Nevertheless, since $S_{1L}(=-S_{L1})$ appears in both (17) and (18), it can be seen that when (17) is satisfied, S_{1L} is relatively small, making the strong reversal of (18) unlikely. The counternormal cases, cannot be ruled out completely, but they are extremely unlikely to prevail.

By adding the income effect term $\partial x^1/\partial I$ to each side of (17), we may transform it into the following uncompensated form,

$$-\frac{\partial x^1}{\partial q^1}/x^1 > \frac{\partial x^1}{\partial q^L}/L \tag{17''}$$

Similarly, (18) may also be so transformed (by adding $\partial L/\partial I$ to each side). From the equivalence of (17), (17′), and (17″), and similarly with respect to (18), it is not difficult to see that Propositions 1 and 2 in Section I are equivalent to Propositions 1′ and 2′, respectively, expressed in different ways. The uncompensated form (17″) has the advantage of being more directly verifiable empirically.

III Concluding Remarks

The propositions established above provide insights into the problem of optimal corrective taxes/subsidies when they have to be financed by distortive taxes. However, to achieve direct policy application, much more work would have to be done. For one thing, the propositions shed light only with respect to the direction, but not the magnitude, of the taxes/subsidies. A useful extension of this analysis is to examine whether the optimal corrective taxes/subsidies should be larger or smaller than those indicated by partial-equilibrium Pigouvian analysis. Secondly, it may be thought that the relaxation of the simplifying assumptions might change the results significantly. Let me examine this briefly.

If the consumer side of the economy cannot be represented by a single utility function, we could either maximize a Paretian social welfare function or maximize one individual's utility subject to the constancy of others. Instead of p^1x^1 in (15), we would then have a weighted sum of various individual consumption of x^1 (for example, $\sum_{k=1}^{s} W_k p^1 x^{1k}$ where k is used to indicate an individual among the s individuals, W_k is his weight, i.e., $w_k \equiv \partial W/\partial U^k$). Similarly, U_E and wL would also be replaced by some appropriate weighted sums. The revised (15′) would then be fairly difficult to handle since both equity and efficiency considerations are blended together. If, for example, a good-producing external diseconomy is predominantly consumed by individuals (the poor?) with high weights, a tax on it may no longer be desirable. However, since we have other instruments (for example, progressive income taxation, transfer payment) to achieve the objective of equity, the separation of equity and efficiency considerations may be justified. (For a further development of this argument, see Appendix 9A of my book.) Moreover, even if the separation of equity and efficiency considerations is not accepted, the case for a tax/subsidy is still valid unless it is overbalanced by the equity consideration caused by, say, a very biased consumption pattern.

REFERENCES

A. B. Atkinson, "How Progressive Should Income-Tax Be?," in Edmund Phelps, ed., *Economic Justice*, Harmondsworth 1973.

______, "Optimal Taxation and the Direct versus Indirect Tax Controversy," *Can. J. Econ.*, Nov. 1977, *10*, 590–606.

J. A. Mirrlees, "An Exploration in the Theory of Optimal Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, 175–208.

______, "Optimal Tax Theory: A Synthesis," *J. Publ. Econ.*, Nov. 1976, *6*, 327–58.

Yew-Kwang Ng, "Non-Economic Activities, Indirect Externalities, and Third-Best Policies," *Kyklos*, No. 3, 1975, *28*, 507–25.

______, *Welfare Economics: Introduction and Development of Basic Concepts*, London 1979.

H. Osana, "Optimal Tax-Subsidy System for an Economy with Marshallian Externalities," *Econometrica*, Mar. 1977, *45*, 329–40.

E. Sadka, "On the Optimal Taxation of Consumption Externalities," *Quart. J. Econ.*, Feb. 1978, *92*, 165–74.

A. Sandmo, "Optimal Taxation in the Presence of Externalities," *Swedish J. Econ.*, Mar. 1975, *77*, 86–98.

______, "Optimal Taxation: An Introduction to the Literature," *J. Publ. Econ.*, June 1976, *6*, 37–54.

The Golden Rule with Endogenous Labor Participation Rate

By LARRY SAMUELSON*

The Golden Rule of accumulation in an economy with an endogenously determined growth rate of labor has been examined by Eric Davis. Davis finds that if the growth rate is an increasing function of per capita net income, steady-state per capita consumption is maximized not at the equality of the propensity to save and capital's share of output, but at a propensity to save smaller than capital's share of output and an interest rate which exceeds rather than equals the growth rate.

The supply of labor in an economy is affected by both the growth and participation rates of its population. Davis' results and the traditional Golden Rule depend upon the implicit assumption that the latter rate is constant. Davis calls for a more flexible treatment of the participation rate, allowing the possibility of optimal unemployment in a theory of optimal savings. This note examines a simple model with a varying participation rate. It is seen that one common assumption describing the participation rate leads to a recommendation opposite that of Davis.

I. The Growth Model

I begin with the basic neoclassical growth model, as in Robert Solow (1956), hereafter called the Solow model. A single homogeneous good Y is produced by labor L, and itself, K, according to the linearly homogeneous production function

$$Y(t)=F[K(t), L(t)]$$

and, letting $y\equiv Y/L$ and $k\equiv K/L$,

$$y\equiv\frac{Y}{L}=F\left[\frac{K}{L},1\right]\equiv f(k)$$

Letting s be the constant propensity to save gives $sf(k)=DK/L$, where D signifies a time derivative, so that $DK=dK/dt$. Assume now that labor grows exponentially at the exogenously given rate n, so that $L(t)=e^{nt}L(0)$ and $DL(t)/L(t)=n$. We then have $Dk=DK/L-kDL/L=DK/L-nk$, and $sf(k)-nk=Dk$. The additional assumptions of $f'(k)>0$, $f''(k)<0$, $f'(0)=\infty$, and $f'(\infty)=0$ are sufficient to guarantee the system will converge to a unique equilibrium at which $sf(k)=nk$, $Dk=0$, and each of the economy's variables grows at rate n.

The Golden Rule identifies the conditions that maximize per capita steady-state consumption c, given by $c=f(k)-nk$. Differentiating this expression with respect to k and setting the result equal to zero gives $f'(k)=n$ and $s=nk/f(k)=kf'(k)/f(k)$. The recommendation is that the interest rate be set equal to the growth rate, and the propensity to save equal to capital's share of income.

The Davis model assumes that the growth rate of the labor force is an increasing function of per capita net income, or $n=n(y)$, with $n'(y)>0$ and $DL(t)=[n(y(t))]L(t)$. Again, the economy converges to a unique equilibrium with $Dk=0$ and $sf(k)=nk$. The conditions for maximum per capita steady-state consumption are found by differentiating $f(k)-nk$ to yield

$$\frac{dc}{dk}=f'(k)[1-kn'(y)]-n=0$$

This gives $f'(k)>n$ and $s=[1-kn'(y)]kf'(k)/f(k)<kf'(k)/f(k)$. The interest rate must be larger than the growth rate, and the propensity to save smaller than capital's share of output. The existence and uniqueness of a solution as well as second-order conditions are examined by Davis.

---

*Assistant professor of economics, Syracuse University.

II. Variable Participation Rate

Let us now return to the realm of an exogenously given growth rate, but allow the participation rate to vary with the wage rate. Letting $J(t)$ be employed labor at time t, we have $J(w,t)=j(w)L(0)e^{nt}$, where $j(w)=J/L$ is the participation rate, with $L(t)=L(0)e^{nt}$ as before, but $Y=F[K(t), J(t)]$. As is often done, assume the supply of labor to be an increasing function of the wage rate, or $j'(w)>0$. To consider the existence of a steady state, now let $k\equiv K/J$, and as before, $sf(k)=DK/J$ and $Dk=DK/J-kDJ/J$. However, we now have

$$\frac{DJ}{J}=n+\frac{j'(w)}{j(w)}Dw$$

Hence

$$sf(k)-nk=Dk+k\frac{j'(w)}{j(w)}Dw$$

Since $w=f(k)-kf'(k)$, we have $w=w(k)$ with $w'(k)=-f''(k)>0$, and this is

$$sf(k)-nk=Dk\left[1+k\frac{j'(w)}{j(w)}w'(k)\right]\cdot$$

The system will converge if the term in brackets is positive, or $1+[kj'(w)w'(k)]/j(w)>0$. This is $(k/j)(dj/dk)>-1$, namely that an elasticity of the participation rate with respect to a capital-labor ratio exceed negative one. This condition allows some cases of $j'(w)<0$, or backward-bending labor supply curves, and is clearly satisfied if $j'(w)>0$. A solution is reached at $sf(k)=nk$ and $Dk=0$.

Again we seek to maximize per capita steady-state consumption. Since the growth rate of the population is fixed, this is equivalent to maximizing total consumption, or $C=F(K,J)-nK$, provided that our attention is restricted to a fixed point in time. The assumption of a steady state ensures that the selection of such a point is arbitrary. Differentiating with respect to the value of capital gives

$$F_K'(K,J)+F_J'(K,J)\frac{dJ}{dK}-n=0$$

or

$$F_K'-n=-F_J'\frac{dJ}{dw}\frac{dw}{dK}$$

From $dw/dK=[w'(k)][dk/dK]=w'(k)[J-K(dJ/dK)]/J^2=[(w'(k))/J]/[1+(dj/dk)(k/j)]$, it is clear that $(dj/dk)(k/j)>-1$, the condition for convergence, ensures $dw/dK>0$. From $j'(w)>0$, we have $dJ/dw>0$, while $F_J'>0$ clearly holds, so that the right-hand side of this equation is negative, and we have a solution of $F_K'=f'(k)<n$ and $s>kf'(k)/f(k)$.[1] The interest rate must fall short of the growth rate and the savings ratio must exceed capital's share of income in order to maximize per capita consumption.[2]

III. A Comparison

To interpret the results of these three models, consider the effect of a change in capital on the value of steady-state per capita consumption. In each model, a change in the value of capital prompts similarly directed changes in per capita net output and the steady-state level of net investment. If the former effect is the greater, an increase in capital boosts net income by an amount sufficient to satisfy the increased investment needs while also contributing to the value of per capita consumption, and the value of per capita steady-state consumption increases as capital does. If the latter effect is the greater, an increase in capital generates less net output than is required to satisfy the needs of net investment, with the balance coming at the expense of consumption. The value of per capita consumption is instead increased by a decrease in capital. The Golden Rule point is found where these two effects are of equal size, and per capita consumption cannot be further increased.[3]

---

[1] The second-order condition is $F_K''+F_{KJ}''(dJ/dK)+(dJ/dK)[F_{JK}''+F_{JJ}''(dJ/dK)]+F_J'(d^2J/dK^2)=F_K''+2F_{JK}''(dJ/dK)+F_{JJ}''(dJ/dK)^2+F_J'[(dJ/dw)(d^2w/dK^2)+(d^2J/dw^2)(dw/dK)^2]$. A maximum is guaranteed if L is not infinite, though the uniqueness of such a maximum requires additional conditions.

[2] These recommendations are obviously reversed for steady states with $j'(w)<0$ that satisfy the convergence conditions.

In the Solow model, the interest rate can be inserted as a proxy for the marginal product of capital, identifying the per unit contribution of an increase in capital to net income, while the growth rate gives the increased obligation to per capita net investment. Their difference signs the derivative of per capita consumption with respect to capital, which is accordingly zero at the Golden Rule point of $r=n$. In the Davis model, the interest rate again describes the marginal product of capital, but the derivative of per capita net investment with respect to capital is $n+kn'(y)f'(k)$, which exceeds n. Here, an increase in capital not only increases the stock upon which net investment must be made, as in the Solow model, but also increases the ratio of net investment to capital stock, so that the per unit obligation to net investment entailed by an increase in capital exceeds the growth rate. Whenever $r=n$, a change in capital affects net investment by a greater amount than net income, and consumption can be increased by employing less capital. The optimal choice is thus an interest rate higher than the growth rate.

The interest and growth rates in the variable participation rate model again identify the marginal product of capital and the unchanging ratio of steady-state net investment to capital. The unique feature of this model is that there is a distinction between the total amount of potential labor in the economy and the amount of labor employed in production, so that the phrase "per capita" must be used with care. Additional insight is gained by considering consumption per labor expended. Letting this quantity be called $\bar{c}$, we have $\bar{c}=Y/J-nK/J$. Differentiating yields

$$\frac{d\bar{c}}{dK}=\frac{J\left[F'_K+F'_J\dfrac{dJ}{dK}\right]-Y\dfrac{dJ}{dK}}{J^2}-n\frac{J-K\dfrac{dJ}{dK}}{J^2}$$

Setting this derivative equal to zero and applying Euler's theorem gives

$$J[F'_K-n]-\frac{dJ}{dK}[F'_K-n]K=0$$

This expression equals zero whenever $F'_K=n$, or whenever $r=n$.

Next, we notice that $c=\bar{c}J/L$, and $dc/dK=(d\bar{c}/dK)J/L+\bar{c}[d(J/L)/dK]$. At $r=n$, the first term of this expression equals zero, while the second is positive. At the equality of interest and growth rates, a marginal increase in capital leaves consumption per expended labor unchanged, while raising the wage rate and hence the participation rate of the population. Per capita consumption is then increased, and is maximized at an interest rate lower than the growth rate. The traditional Golden Rule point of $r=n$ maximizes not per capita consumption but consumption per expended labor,[4] and because it does so, per capita consumption is maximized by setting $r<n$.

IV. Conclusion

The result of the variable participation rate model is a simple theory of optimal steady-state savings and employment. Interest rates greater than the growth rate are clearly not optimal, since a decrease in the interest rate offers a higher participation rate and a higher output of consumption goods per labor expended. The region of interest rates smaller than the growth rate offers the economy a choice. An increase in the participation rate, and hence employment, is accomplished only at the cost of decreased production of consumption goods per labor expended, while an increase in consumption output per labor expended is accompanied by a decreasing level of employment. Per capita consumption is maximized by balancing these two effects. It is optimal to withhold part of the potential labor force from employment, thereby boosting the consumption goods output of existing labor enough to increase per capita consumption. At the same time, it is optimal to employ more workers than would be employed at the traditional Golden Rule point of $r=n$, with the decreased output of consumption per expended labor being more than offset by the increasing participation rate.

---

[3] Davis offers the statement that an increase in the savings ratio reduces current consumption while increasing future consumption, and that these are the two effects balanced at the Golden Rule point. This could be true if we were examining the actual transition from one steady state to another, but is not the situation addressed by the Golden Rule. Instead, the Golden Rule asks which of many different steady-state economies offers the highest constant level of steady-state consumption, meaning the highest level of both present and future consumption. The hypothetical jump from one such economy to another either increases or decreases consumption in all periods, present and future, and it is the forces tending to make this change in consumption an increase or decrease that are balanced by the Golden Rule.

[4] Davis notes that if the growth rate is an increasing function of per capita consumption, the latter is maximized at $r=n$. A similar result obtains in our model if the participation rate is an increasing function of consumption per employed labor.

The model is curious in that if the participation rate is increasing in the wage rate, its prescription is the opposite of Davis', calling for an interest rate lower instead of higher than the growth rate. If we make the heroic assumption that these highly simplified models are not so removed from reality as to be devoid of practical value, we obtain divergent assessments of present economic experience, with the Davis model suggesting those economies in which the interest rate exceeds the growth rate are not as divergent from optimality as the traditional Golden Rule would imply, while the variable participation rate model identifies them as being even further from optimality than indicated by the original Golden Rule. A deeper analysis of this question calls for more flexible models.

A first extension of these models could be the combination of both a variable growth and a participation rate. Other kinds of unemployment may also be desired, as well as the ability to examine economies which are not in steady states. Along the way, the reaction of the growth rate to changes in net income and the participation rate to variations in the wage rate must be examined. If the participation rate is decreasing rather than increasing in the wage rate, my model would agree with Davis' in prescribing $r>n$. At the same time, if the growth rate slows rather than quickens as income increases, Davis' model would agree with mine in prescribing $r<n$. Neither of these appears to be a completely settled question.

REFERENCES

E. Davis, "A Modified Golden Rule: The Case With Endogenous Labor Supply," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 177–81.

E. Phelps, "The Golden Rule of Accumulation: A Fable for Growthmen," *Amer. Econ. Rev.*, Sept. 1961, *51*, 638–43.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 65–94.

______, "Comment on the Golden Rule," *Rev. Econ. Stud.*, June 1962, *29*, 255–57.

Incentives and the Choice of Optimal Plans

By SUSAN I. COHEN*

There is a vast literature in economics concerned with the specification and discussion of planning procedures for economies (for example, see Kenneth Arrow; Theodore Groves; Geoffrey Heal; Martin Loeb and Wesley Magat). This literature is generally concerned with the algorithmic side of planning, in that it deals with behavioral rules that if followed lead to a "social optimum" (from the planner's point of view). In general, the enforcement question is not dealt with; namely, is it in the best interests of the economic agents to follow the planning board's directives?[1]

If it is not in the economic agents' interests to follow the planning board's directives, then it is not necessarily true that a social optimum will actually be attained by these procedures. The existence of enforcement rules or evaluation measures that insure compliance has been shown in several rather general contexts (see, for example, Groves). The importance of the existence of such rules is well recognized in the planning literature:

> There is a very important aspect of any planning problem...[that] may be loosely described as the problem of ensuring that it is in the interests of the firms in the economy to behave as the centre requires them to.... Once an incentive structure has been specified for a planned economy, it is reasonable to suppose that firms will be concerned to choose that strategy that will maximise the rewards accruing to them under this.... Ideally one would like to devise for any planning process an associated incentive structure such that each firm would maximise its returns under this system by revealing correct information to the centre, and producing the optimum once located.
> [Heal, p. 214]

The purpose of this paper is the definition of just such a system for a particular class of decentralized planning procedures (and thus a particular model of a planned economy). As noted above, the existence of such a system has been previously demonstrated in other contexts. However, those results are not applicable to decentralized planning procedures since those systems require firms to communicate entire functions or sets as a one-time message. The properties of those procedures depend heavily on the assumption of noniterative communication. A general model with iterative communication is presented in my earlier paper. The results obtained there are weaker than those obtained when communication is noniterative. In particular, Groves demonstrates that when communication is noniterative, following the planning board's directives is a dominant strategy for every firm.[2] When communication is iterative, it is only possible to show that following the planning board's directives constitutes a Nash equilibrium that weakly dominates all other Nash equilibria for all firms.

In Section I, a model of an economy (based on a model due to Edmond Malinvaud) is presented. A class of planning procedures for the economy is defined in Section II. The enforcement rules for these procedures are defined in Section III, where it is also shown that, with these rules, the agents in the economy have an incentive to follow the planning board's directives. Finally, in Section IV, it is shown that Malinvaud's model and "Procedure Implying Mathematical Programming at the Center" are special cases of the model and planning procedures defined in Section II.

---

*Assistant professor, College of Industrial Management, Georgia Institute of Technology. This research was partially supported by National Science Foundation Grant SOC 78-05770. I would like to thank an anonymous referee for very helpful suggestions. All errors, of course, remain my own.

[1] Jacques Dreze and D. de la Vallee Poussin, John Roberts, and Francoise Schoumaker consider these questions in tâtonnement procedures. Loeb and Magat consider the problem in a centrally planned economy in which the communication between the planing board and the firms is of a "one-time only" nature.

[2] The models in my earlier paper and in Groves are not stated as planning models, although they can easily be interpreted as such.

I. The Model of the Economy

Let us consider the following model[3] of an economy: there are $N+1$ agents, consisting of N firms (or ministries) and the planning board. There are M commodities produced in the economy. The net demand for each good produced is the final consumption of the good minus the net outputs of the good by all firms. A program P for the economy consists of a vector of final consumption, a vector of net outputs for all goods for all firms, and a vector of net demands for all goods.

The program P chosen by the planning bureau must be feasible from the consumers' point of view. Let us assume that the planning bureau knows, a priori, the set of feasible final consumption vectors. The program P must also be technologically feasible; thus the vector of net outputs for each firm must be an element of the firm's feasible production set. Each firm knows[4] its own production-possibilities set. These sets are not known to the planning bureau. Finally, the net demand for each good in the program chosen by the planning bureau must be feasible, that is, the net demand cannot exceed the initial amount of the good available in the economy (which is known by the planning bureau).

Social preferences among alternative vectors of final consumption are represented by a utility function, known by the planning bureau. A program P^1 is preferred to P^2 if the utility of final consumption in P^1 is higher than the utility of final consumption in P^2. The planning bureau seeks a feasible program in which the utility of final consumption is greater than or equal to the utility of final consumption in all other feasible programs.

Formally, the model of the economy is as follows:

(a) x_k is the final consumption of good k, $k=1,\ldots,M$. Let $x\equiv(x_1,\ldots,x_M)$.

(b) y_{ki} is the net output of good k by firm i, $k=1,\ldots,M$, $i=1,\ldots,N$.

$y^k\equiv\sum_{i=1}^{N} y_{ki}$ is the total output of good k.

$y_i\equiv(y_{1i},\ldots,y_{Mi})$ is the vector of net outputs of all goods by firm i. Let $y\equiv(y_1,\ldots,y_N)$.

(c) Y_i is firm i's feasible production set, $i=1,\ldots,N$. W_k is the initial amount of good k available in the economy, $k=1,\ldots,M$.

(d) X is the set of feasible final consumption vectors.

(e) $U(x)$ is the utility of the final consumption x. $U(\cdot)$ is a function from the commodity space $\mathbb{R}^M$ (the M-fold Cartesian product of the real line) into the real line; i.e., $U\colon \mathbb{R}^M\rightarrow \mathbb{R}$.

The planning bureau seeks the program $P^*=(x^*, y^*)$ that solves the planning problem:

$$\max_{x,y} U(x) \text{ subject to } x\in X$$

$$y_i\in Y_i \quad i=1,\ldots,N \tag{1}$$

$$x-\sum_{i=1}^{N} y_i\leqslant W$$

II. A Class of Planning Procedures

The class of planning procedures I consider are those that choose feasible solutions and coverage to an optimal plan (either finitely or in the limit). I therefore eliminate from consideration those procedures that ensure ε-convergence, that is, a plan "close" to the optimal plan.

To reiterate, let us assume that the planning board knows its own utility function $U(x)$ as well as the set of feasible final consumption vectors X and the vector W of initial availabilities of all goods. Each firm knows its own feasible production set Y_i, but not the sets of the other firms; the firms do not know the utility function of the planning board nor the set X. If the planning board knew all the production sets Y_i, it could solve its planning problem (1) directly.

---

[3] This model is the one developed by Malinvaud.

[4] By "know" I mean that the firm has the capability of calculating the coordinates of any point in its production set; for example, given a price vector, it can determine the vector in its production set with maximum value.

The purpose of the planning procedure is thus to obtain information from the firms about their feasible production sets. We will assume that the planning board knows that the Y_i are all members of the same subset $\mathbf{H}$ of sets in the commodity space $\mathbb{R}^M$. For example, $\mathbf{H}$ may be the set of convex subsets of $\mathbb{R}^M$, so that the planning board knows that all the firms' production-possibilities sets are convex. If the firms choose to misrepresent their production sets (i.e., cheat), they must do so in a way that is not detectable by the planning board.

More formally, if Y_i is firm i's true production set, let $\mathbf{Y}_i \subset \mathbb{R}^M$ denote firm i's *reported* production set. Firm i will be said to be "telling the truth" if its reported production set is its true production, i.e., if $\mathbf{Y}_i = Y_i$.

At each stage or iteration of the planning process, the planning bureau receives messages from the firms in response to its messages. The set of possible firm messages and planning board messages are denoted by $\mathbf{M}$ and $\mathbf{P}$, respectively. For example, the firms might be reporting production vectors, so that $\mathbf{M} = \mathbb{R}^M$, while the planning board might be sending price vectors, so that $\mathbf{P}$ is the unit simplex in $\mathbb{R}^M$. Let $p_i^t \in \mathbf{P}$ be the message sent to firm i at iteration t; let $m_i^t \in \mathbf{M}$ be the reply of firm i.

DEFINITION 1: *The sequence* $\{m_i^t\}_{t=1}^{\infty}$ *will be said to be consistent with a set* $\mathbf{Y}_i$ *given* $\{p_i^t\}_{t=1}^{\infty}$ *if* $\{m_i^t\}_{i=1}^{\infty}$ *are the messages that would result from following the planning board's directives when* $\mathbf{Y}_i \subset \mathbb{R}^M$ *is the firm's production set.*[5]

Thus, a sequence of messages $\{m_i^t\}_{t=1}^{\infty}$ is consistent with a set in $\mathbf{H}$ if $\mathbf{Y}_i \in \mathbf{H}$ exists satisfying Definition 1.

The messages sent to the planning board by each firm depend upon the production set $\mathbf{Y}_i$ the firm chooses to report as well as the board's messages $\{p_i^t\}_{t=1}^{\infty}$. The planning board's messages to each division are assumed to be nonrandom, that is, functionally dependent on the reported production sets of *all* the firms. We can express each firm's messages as a function of *all* the reported production sets:

$$m_i^t = m_i^t(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$$

All the messages received by the planning board from every firm at every iteration will be denoted by $m^{\infty} \equiv (m_1^1, \ldots, m_N^1, m_1^2, \ldots, m_N^2, \ldots)$.

Define $(x(m^{\infty}), y(m^{\infty}))$ to be the vector of final consumption and firm production plans chosen by the planning procedure when the firms send messages m^{∞} to the planning board:

DEFINITION 2: *A planning procedure will be called admissible if when the reported production sets* $\mathbf{Y}_i \in \mathbf{H}$ *for* $i = 1, \ldots, N$, $(x(m^{\infty}(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)), y(m^{\infty}(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)))$ *solves the planning problem* (1) *where* Y_i *has been replaced by* $\mathbf{Y}_i$ *for all* i.

Given this definition of an admissible procedure, if the firms all report their true production sets Y_i, the procedure finds the solution to the planning board's problem (1).

III. Definition of Firm Evaluations

Given the definition of an admissible planning procedure, we can define the enforcement rules. The enforcement rules are defined implicitly by an N-tuple of evaluation measures, one for each firm. When the evaluation measures are appropriately defined, maximization by a firm of its individual measure "enforces" the following of the planning board's directives. It should be pointed out that these evaluations provide index numbers on which individual firm rewards can be based; they do not represent real resource flows. For example, the reward to one firm might be in the form of paid vacation days, while another firm might be rewarded with salary increases.

---

[5] Since it may not be decidable in general whether or not $\{m_i^t\}_{t=1}^{\infty}$ is consistent with a particular set $\mathbf{Y}_i \subset \mathbb{R}^M$, I am implicitly excluding planning procedures for which this is not decidable.

DEFINITION 3: *Firm i's evaluation measure is a function* $E_i: \mathbf{M}^{\infty} \to \mathbb{R}$ *defined as follows:*[6]

$$E_i[m^{\infty}] \equiv \begin{cases} U(x(m^{\infty})) \text{ if } \{m_i^t\}_{t=1}^{\infty} \text{ is consistent with a set in } \mathbf{H} \text{ given } \{p_i^t\}_{t=1}^{\infty} \text{ and } y_i(m^{\infty}) \text{ is feasible} \\ \inf_{x \in X} U(x) - 1 \text{ otherwise} \end{cases}$$

These evaluations do not preclude firm cheating directly. The firm must appear to be reporting its true production set; in addition if $y_i(m^{\infty}) \notin Y_i$ and thus not implementable, the planning board knows with certainty that the firm lied about its production set.

Payoff to each firm can be expressed as a function of the reported production sets:

$$W_i[\mathbf{Y}_1,\ldots,\mathbf{Y}_N] \equiv E_i[m^{\infty}(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)]$$

The problem for each firm is thus the choice of a reported production set that maximizes the firm's payoff.

If firm i knows only that the other firms are reporting production sets in **H** (possibly the true production sets, but not necessarily), firm i's best response is to report its true production set Y_i.

Formal (mathematical) statements of all theorems and corollaries as well as their proofs (when necessary) are given in the Appendix. In addition, although not stated in each theorem and corollary explicitly, the results are only shown to hold for the planning procedures and evaluation measures defined by Definitions 2 and 3, respectively.

THEOREM 1: *Firm i maximizes its payoff by reporting its true production set if the other* $(N-1)$ *firms all report production sets in the set* **H**.

The following is immediate from Theorem 1:

COROLLARY 1: *Firm i maximizes its payoff by reporting its true production set when every other firm is reporting its true production set; i.e.,* $(Y_1,\ldots,Y_N)$ *is a Nash equilibrium.*

Is telling the truth the only optimal strategy for firm i? The answer is no. Nevertheless, one can show that any production set $\mathbf{Y}_i^*$ that maximizes firm i's payoff for any possible (reported) production sets in **H** of the other firms must have a nonempty intersection with Y_i, i.e., $\mathbf{Y}_i^* \cap Y_i \neq \emptyset$.

THEOREM 2: *If firm i maximizes its payoff by reporting* $\mathbf{Y}_i^*$ *no matter what production sets in* **H** *the other* $(N-1)$ *firms report, then* $\mathbf{Y}_i^* \cap Y_i \neq \emptyset$.

Firm i will cheat only if cheating cannot be detected *and* does not reduce the firm's payoff if undetected. Therefore firm i will exclude points of its true production set only if there exist no production sets of the other firms that would lead the planning bureau to choose the excluded point given the set X and its constraints on feasible final consumption. Similarly, firm i will include infeasible points in its reported production set only if it is sure these infeasible points will never be chosen.

Calculation of an optimal production set to report (other than Y_i) requires knowledge of the set of feasible final consumption vectors X. Thus Y_i is the only reported production set that maximizes payoff and is totally independent of the reported production sets of the other firms, and the set X.

Intermixtures of Nash equilibria are generally not Nash equilibria; we would thus like to ensure that the firms have an incentive to arrive at a Nash equilibrium that leads to a decision that solves problem (1). The N-tuple $(Y_1,\ldots,Y_N)$ can be shown to dominate for every firm in the set of Nash

[6] $\mathbf{M}^{\infty} \equiv \mathop{X}_{t=1}^{\infty} \mathbf{M}$

equilibria; in addition, any such dominant Nash equilibrium leads to a decision solving problem (1).

THEOREM 3: *The payoff to each firm is at least as great at the Nash equilibrium* $(Y_1,\dots,Y_N)$ *as it is at any other Nash equilibrium.*

THEOREM 4: *If the payoff to each firm under the Nash equilibrium* $(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)$ *is the same as under the Nash equilibrium* $(Y_1,\dots,Y_N)$, *i.e., when every firm reports its true production set, then*

$$\left(x(m^\infty(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)),y(m^\infty(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*))\right)$$

solves the planning board's problem (1).

IV. A Procedure Implying Mathematical Programming at the Center (*MPC*)

Malinvaud defines three planning procedures for solving problem (1). The first is a tâtonnement procedure; the second is a procedure based on the Leontief-Samuelson technology; and the third is a procedure implying the use of mathematical programming by the planning bureau. It is this third procedure with which this section is concerned. Malinvaud makes the following assumptions about $U(\cdot)$, X, the sets Y_i, and the planning board's a priori knowledge:

(a) The set X is closed and convex. There exists an $\bar{x}\in X$ such that $x\geqslant\bar{x}$ for all $x\in X$. The utility function $U(x)$ is concave and continuous.

(b) The sets Y_i, $i=1,\dots,N$, are closed, bounded and convex.

(c) The planning board knows a feasible solution to the planning problem (1).

In this procedure prices are used to decompose problem (1). At each stage of the process, the planning bureau sends the firm a (normalized) price vector p^{t-1} for their outputs. Each firm then maximizes the value of its output at the price vector, that is, each firm solves a problem of the following form:

$$\text{(2)}\qquad \max_{y_i} p^{t-1}y_i \quad \text{subject to } y_i\in Y_i$$

The firm then sends to the planning bureau the solution y_i^{t-1} to the firm's problem. The bureau uses this maximizer along with the maximizers received at previous iterations to approximate Y_i, the firm's feasible production set. The approximated production set Y_i^t for firm i at iteration t is the set of all convex combinations of maximizers received from firm i at all iterations, including the current one:

$$Y_i^t\equiv\left\{y_i \,\middle|\, y_i=\sum_{\tau=1}^{t-1}\lambda_{i\tau}y_i^\tau,\ \sum_{\tau=0}^{t-1}\lambda_{i\tau}=1,\ \lambda_{i\tau}\geqslant 0,\ \tau=0,1,\dots,t-1\right\}$$

The planning bureau then solves for what the optimal vectors of final consumption and net outputs would be if the firms' production sets were actually the sets Y_i^t, $i=1,\dots,N$, i.e., the planning bureau solves problem (1) with Y_i replaced by Y_i^t for all i. Let $(x^t,y^{(t)})$ solve the planning board's problem at iteration t. Revised prices p^t are any (normalized) M-vector such that $p^t\geqslant 0$ and satisfying the following three conditions:

1) $p^tx\succ p^tx^t$ for all $x\in X$ such that $U(x)\succ U(x^t)$

2) $p^ty_i^{(t)}\leqslant p^ty_i$ for all $y_i\in Y_i^t$, for all $i=1,\dots,N$

3) $p_k^t=0$ for all k such that

$$x_k^t-\sum_{i=1}^{N}y_{ki}^{(t)}\prec W_k$$

The planning bureau's decision is

$$(x(m^\infty),y(m^\infty))\equiv\left(\lim_{t\to\infty}x^t,\ \lim_{t\to\infty}y^{(t)}\right)$$

Since the planning procedure converges to a solution of problem (1) (see Malinvaud) it is straightforward to verify its admissibility. If at each stage each division i solves a problem like (2) where Y_i is replaced by $\mathbf{Y}_i$, and where $\mathbf{Y}_i$ is convex, then it is immediate by admissibility that

$$\left(x(m^\infty(\mathbf{Y}_1,\dots,\mathbf{Y}_N)),y(m^\infty(\mathbf{Y}_1,\dots,\mathbf{Y}_N))\right)$$

solves problem (1) where Y_i has been replaced by $\mathbf{Y}_i$ for all i.

In order to apply the evaluation measures defined in Section III, and therefore the results of that section, it must be shown that the planning bureau can judge whether or not a sequence of vectors $\{y_i^t\}_{t=1}^{\infty}$ is consistent with some convex set $\mathbf{Y}_i$. Since the firm is maximizing a linear function over a convex set, the maximizer y_i must be an extreme point of $\mathbf{Y}_i$. Therefore a sequence $\{y_i^t\}_{t=1}^{\infty}$ will be consistent with a convex set $\mathbf{Y}_i$ if and only if the following is true for all t:

$$y_i^t \notin (\text{interior } Y_i^t)$$

If $\mathbf{Y}_i$ were a convex polyhedral set, then y_i^t could certainly be a point on one of the faces of Y_i^t; i.e., $y_i^t \in (\text{boundary } Y_i^t)$.

The evaluation measure for each firm is then:

$$E_i[m^{\infty}] = \begin{cases} U(x(m^{\infty})) & \text{if } y_i^t \notin (\text{interior } Y_i^t) \\ & \text{for all } t; \\ & \text{and } \lim\limits_{t \to \infty} y_i^{(t)} \in Y_i \\ \inf\limits_{x \in X} U(x) - 1 & \text{otherwise} \end{cases}$$

If the firms in Malinvaud's economy are evaluated in the above manner, there is reason to expect that the rules of *MPC* will be followed, based on the results in Section III. These results state that the N-tuple of true production sets $(Y_1, \dots, Y_N)$ constitute a Nash equilibrium that dominates in the set of Nash equilibria. Moreover, this is the only Nash equilibrium known a priori to all the firms in the economy and that is independent of the set X of feasible final consumption vectors.

APPENDIX

All theorems and corollaries hold for all admissible planning procedures and evaluation measures defined by Definitions 2 and 3, respectively.

THEOREM 1: *For all* $\mathbf{Y}_j \in \mathbf{H}, j \neq i$, *if* $\mathbf{Y}_i^* = Y_i$ *then* $W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_i^*, \dots, \mathbf{Y}_N] \geqslant W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N]$ *for all* $\mathbf{Y}_i \subset \mathbb{R}^M$.

PROOF:

By admissibility of the planning procedure, $(x(m^{\infty}(\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N)), y(m^{\infty}(\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N)))$ solves

$$\max_{x,y} U(x)$$

subject to $x \in X$; $y_i \in Y_i$; $y_j \in \mathbf{Y}_j$ $j \neq i$; and $x - \sum_{i=1}^{N} y_i \leqslant W$.

Let $\mathbf{Y}_i$ be any other subset of $\mathbb{R}^M$ (not necessarily convex), and $(x(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_N)), y(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_N)))$ the vector of final consumption and production chosen by the planning board when firm i reports its production set as $\mathbf{Y}_i$. One of the following must hold:

(i) $y_i(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_N)) \in Y_i$. The production plan chosen when firm i reports $\mathbf{Y}_i$ is still available if the firm reports Y_i, in addition to production plans that were not previously available. It is immediate that $U(x(m^{\infty}(\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N))) \succcurlyeq U(x(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_N)))$ and $W_i[\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N] \succcurlyeq W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N]$.

(ii) $y_i(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_N)) \notin Y_i$. Then $W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N] = \inf_{x \in X} U(x) - 1$ and $W_i[\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N] \succ W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N]$.

COROLLARY 1: *The* N*-tuple* $(Y_1, \dots, Y_N)$ *is a Nash equilibrium, i.e., for all* $i = 1, \dots, N$, $W_i[Y_1, \dots, Y_N] \geqslant W_i[Y_1, \dots, \mathbf{Y}_i, \dots, Y_N]$ *for all* $\mathbf{Y}_i \subset \mathbb{R}^M$.

PROOF:

Immediate from Theorem 1.

THEOREM 2: *If for all* $\mathbf{Y}_j \in \mathbf{H}$, $j \neq i$, $W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_i^*, \dots, \mathbf{Y}_N] \geqslant W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N]$ *for all* $\mathbf{Y}_i \subset \mathbb{R}^M$, *then* $\mathbf{Y}_i^* \cap Y_i \neq \phi$.

PROOF:

Assume $\mathbf{Y}_i^* \cap Y_i = \phi$. Then $y_i(m^{\infty}(\mathbf{Y}_1, \dots, \mathbf{Y}_i^*, \dots; \mathbf{Y}_N) \notin Y_i$ for all convex $\mathbf{Y}_j$, $j \neq i$, and $W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_i^*, \dots, \mathbf{Y}_N] = \inf_{x \in X} U(x) - 1$ for some convex $\mathbf{Y}_j$, $j \neq i$. Therefore $W_i[\mathbf{Y}_1, \dots, Y_i, \dots, \mathbf{Y}_N] \succ W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_i^*, \dots, \mathbf{Y}_N]$ which is a contradiction (from Theorem 1).

THEOREM 3: *For every Nash equilibrium* $(\mathbf{Y}_1, \dots, \mathbf{Y}_N)$, $W_i[Y_1, \dots, Y_N] \geqslant W_i[\mathbf{Y}_1, \dots, \mathbf{Y}_N]$ *for all* $i = 1, \dots, N$.

PROOF:

If $P=(x(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N)),\ y(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N)))$ is chosen when the firms report $(\mathbf{Y}_1,\dots,\mathbf{Y}_N)$, when the firms report $\mathbf{Y}_i'=\{y_i(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N))\}$, the planning board will again choose P. Therefore, $U(x(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N)))=U(x(m^{\infty}(\mathbf{Y}_1',\dots,\mathbf{Y}_N')))$. Since $(\mathbf{Y}_1,\dots,\mathbf{Y}_N)$ is a Nash equilibrium, $y_i(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N))\in Y_i$ for all i. Thus, $\mathbf{Y}_i'\subset Y_i$ for all i and $U(x(m^{\infty}(\mathbf{Y}_1',\dots,\mathbf{Y}_N')))\leqslant U(x(m^{\infty}(Y_i,\dots,Y_N)))$.

We therefore have $W_i[\mathbf{Y}_1,\dots,\mathbf{Y}_N]\equiv U(x(m^{\infty}(\mathbf{Y}_1,\dots,\mathbf{Y}_N)))=U(x(m^{\infty}(\mathbf{Y}_1',\dots,\mathbf{Y}_N')))\leqslant U(x(m^{\infty}(Y_1,\dots,Y_N)))\equiv W_i[Y_1,\dots,Y_N]$

THEOREM 4: *If* $(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)$ *is a Nash equilibrium and for all* i, $W_i[Y_1,\dots,Y_N]=W_i[\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*]$, *then* $P^*=(x(m^{\infty}(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)),\ y(m^{\infty}(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)))$ *solves problem* (1).

PROOF:

(i) Since $(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)$ is a Nash equilibrium, $P^*=(x(m^{\infty}(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)),\ Y(m^{\infty}(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)))\in X\times Y_1\times\dots\times Y_N$

(ii) By admissibility, $(x(m^{\infty}(Y_1,\dots,Y_N)),\ y(m^{\infty}(Y_1,\dots,Y_N)))$ solves (1). By assumption, $U(x(m^{\infty}(Y_1,\dots,Y_N)))\equiv W_i[Y_1,\dots,Y_N]=W_i[\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*]\equiv U(x(m^{\infty}(\mathbf{Y}_1^*,\dots,\mathbf{Y}_N^*)))$. From (i), P^* is feasible. Thus, P^* solves (1).

REFERENCES

K. Arrow, "Control in Large Organizations," *Manage. Sci.*, Apr. 1964, *10*, 397–408.

______ **and L. Hurwicz**, "Decentralization and Computation in Resource Allocation," in Ralph Pfouts, ed., *Essays in Economics and Econometrics*, Chapel Hill 1960, 34–104.

S. Cohen, "Incentives, Iterative Communication and Organizational Control," *J. Econ. Theory*, Feb. 1980, *22*, 37–55..

______, "Incentive Compatible Control of the Multidivisional Firm with Iterative Communication," unpublished doctoral dissertation, Northwestern Univ. 1977.

J. Drèze and D. de la Vallee Poussin, "A Tâtonnement Process for Public Goods," *Rev. Econ. Stud.*, Apr. 1971, *38*, 133–50.

T. Groves, "Incentive Compatible Control of Decentralized Organizations," work. paper no. 166, Center Math. Stud. Econ. Manage. Sci., Northwestern Univ., Aug. 1975.

Geoffrey Heal, *The Theory of Economic Planning*, Amsterdam 1973.

L. Hurwicz, "The Design of Mechanisms for Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 1–30.

P. Jennergren, "Studies in the Mathematical Theory of Decentralized Resource Allocation," unpublished doctoral dissertation, Stanford Univ. 1971.

M. Loeb and W. Magat, "Success Indicators in the Soviet Union: The Problem of Incentives and Efficient Allocation," *Amer. Econ. Rev.*, Mar. 1978, *68*, 173–81.

E. Malinvaud, "Decentralized Procedures for Planning," in his and M. O. L. Bacharach, eds, *Activity Analysis in the Theory of Growth and Planning*, London 1967, 170–208.

J. Roberts, "Incentives in Planning Procedures for the Provision of Public Goods," *Rev. Econ. Stud.*, forthcoming.

F. Schoumaker, "Best Reply Strategies in the Champsaur-Drèze-Henry Process," unpublished manuscript, Northwestern Univ., Nov. 1977.

The Forward Exchange Rate, Expectations, and the Demand for Money— The German Hyperinflation: Comment

By MICHAEL K. SALEMI*

As Phillip Cagan pointed out, hyperinflation provides the opportunity to study monetary phenomena in a situation where increases in nominal quantities dwarf changes in real quantities. The large changes in prices, moreover, surely create sizeable incentives to predict as well as possible the changes in these prices. Thus a central feature of a model of asset supply and demand in hyperinflation is the way in which agents are assumed to form their expectations. In a recent article in this *Review*, Jacob Frenkel proposed to infer from the data on spot and forward deutsche mark exchange the one-month future rate of inflation expected in Germany during the post-World War I hyperinflation. The major virtue of such an approach, suggested Frenkel, was that it depended on observable market prices rather than mechanistic formulae to generate agents' guesses as to the opportunity cost of holding money.

Frenkel based his conclusion that foreign exchange data could be used to measure inflation expectations on evidence that during the hyperinflation the market in deutsche mark exchange functioned efficiently. He based his conclusion that exchange markets were efficient on his estimates of two regressions:

$$(1)\quad \log S_t = a + b \log F_{t-1} + U_t$$

$$(2)\quad \log S_t = a' + b' \log F_{t-1} + c' \log F_{t-2} + U_t'$$

where S_t is the spot rate between German marks and pounds sterling, and F_{t-j} is the one-month forward rate at time $t-j$, that is, the price at time $t-j$ for delivery at $t-j+1$. If the foreign exchange market is efficient, Frenkel argued, one would expect to find that the constants are zero, the coefficients of F_{t-1} are one, the coefficient of F_{t-2} is zero, and that the errors are serially uncorrelated. On the basis of his estimation of (1) and (2), Frenkel found that these restrictions were at least approximately met and concluded that the data from the foreign exchange market was appropriately used to measure expectations.[1]

Evidence will be presented below that the market in deutsche mark exchange did *not* function efficiently. In particular it will be argued that the forward exchange premium did not contain all the information available to agents to predict the rate of inflation. Thus the view that inflation expectations based solely on the forward exchange premium were rational in the sense of John Muth will be rejected.

I. Inflation Expectations and the Forward Exchange Premium

That the forward exchange premium should contain information on the rate of inflation expected to be realized one period

---

*Assistant professor of economics, University of North Carolina-Chapel Hill. I wish to acknowledge helpful comments from Arthur Benavie, Richard Froyen, James Ingram, Thomas Sargent, Roger Waud, and the managing editor. Part of the work of this paper was assisted financially by the Social Science Research Council.

[1] "The results," concludes Frenkel, "...provide support for the notion that during the hyperinflation expectations may have behaved rationally in the sense of Muth, and that one may use data from the foreign exchange market to measure expectations. In fact, it should not be surprising that even during the turbulent period of the hyperinflation the rational expectations hypothesis might apply. It stands to reason that the larger variability of the exchange rate increases the rate of return from and the amount of resources invested in accurate forecasting" (pp. 655–56).

in the future is a consequence of the expectations hypothesis. The hypothesis states that agents will be indifferent between prospects if and only if their expected returns are equal, and implies

$$(3) \qquad (1+r_{uk,t})=\frac{S_t}{F_t}(1+r_{d,t})$$

where $S_t \equiv$ the spot rate of exchange at time t (DM/£)

$F_t \equiv$ the price at t of exchange to be delivered at t+1

$r_{uk,t} \equiv$ short-term bank rate in Britain

$r_{d,t} \equiv$ short-term bank rate in Germany.[2]

The left-hand side of (3) is the return to a British wealth owner holding short-term British bonds. The right-hand side of (3) is the return to the same wealth owner who holds short-term German bonds and adopts a covered position in foreign exchange.

Equation (3) itself does not imply that agents build into the foreign exchange premium all the information in economic data useful for predicting the inflation rate. But suppose further that

$$(4) \qquad r_{d,t}=\pi_{d,t}+\bar{r}_d$$

$$r_{uk,t}=\pi_{uk,t}+\bar{r}_{uk}$$

where πs denote rates of inflation anticipated at t to be realized at t+1 and $\bar{r}$s denote real rates of interest. By substituting (4) into (3) and taking *logs* one obtains

$$(5) \qquad ln(1+\pi_{uk,t}+\bar{r}_{uk}) = ln(S_t/F_t)+ln(1+\pi_{d,t}+\bar{r}_d)$$

which implies approximately

$$(6) \qquad \pi_{d,t} \cong ln(F_t/S_t)+\pi_{uk,t}+\bar{r}_{uk}-\bar{r}_d$$

Since $ln\,(F_t/S_t)$ approximately equals $(F_t - S_t)/S_t$, the premium for forward exchange in percentage terms, equation (6) forms the basis for linking $\pi_{d,t}$ to the forward premium. If the real rate in the United Kingdom is equal to the German real rate, (6) predicts that the expected inflation rate of German prices will be approximately equal to the forward premium plus the expected inflation rate of British prices.

Observations on British nominal interest rates during the period of the hyperinflation are available from several sources. In what follows the Frenkel inflation predictor will be taken to be

$$(7) \qquad \pi_{d,t} \cong FP_t-\bar{r}_d$$

$$\text{where} \qquad FP_t \equiv ln(F_t/S_t)+r_{uk,t}$$

The effect of $\bar{r}_d$ on the prediction will be accounted for by the inclusion of a constant in the equations which are estimated. Note that (7) implies that if $\bar{r}_d$ is positive then the forward premium should *exceed* $\pi_{d,t}$.

II. Testing the Rationality of Inflation Predictions Based on the Forward Exchange Premium

Frenkel's line of investigation, while important and interesting in any case, would clearly be more interesting if his conjecture on the rationality of inflation predictions based on the forward exchange premium were true. A direct test of the conjecture is available, and suggests, however, that the conjecture is false. The test depends in a straightforward way on the fact that rationally formed expectations do not waste information which would be useful in forming a prediction. Thomas Sargent and Neil

---

[2] The data for the deutsche mark spot and forward exchange are from Paul Einzig which gives spot and forward quotes at weekly intervals. For this study the quote nearest the fifteenth of each month was employed. A second set of data for forward and spot exchange in Germany are reported in John Maynard Keynes. The data in Keynes are for the beginning of the month. The interest rate employed in this study is the short-term bank rate in London reported in the *Federal Reserve Bulletin* (1921, 1922, 1923) and elsewhere. The rate of inflation X_t is the first difference of the natural *log* of the index of wholesale prices P_t observed as a monthly average of daily rates throughout the hyperinflation and reported in Statistisches Reichsamt.

TABLE 1—TEST OF THE RATIONALITY OF THE FORWARD EXCHANGE PREMIUM PREDICTION OF THE RATE OF INFLATION

	Dependent Variable	con.	*trend*	FP_{t-1}	X_{t-1}	X_{t-2}	X_{t-3}	X_{-4}	X_{t-5}	X_{t-6}	
Feb. 1921–	X_t	.18	–	5.21	–	–	–	–	–	–	$R^2=.43$
Aug. 1923		(.08)		(1.11)							$D.W.=1.33$
											$SSR=4.71$
	X_t	−.72	.04	2.36	.62	−.54	.28	−1.01	−.40	−.70	$R^2=.80$
		(.29)	(0.1)	(1.49)	(.28)	(.34)	(.36)	(.36)	(.35)	(.41)	$SSR=1.67$
	X_t-FP_{t-1}	.32	–	–	–	–	–	–	–	–	$D.W.=.52$
		(.09)									$SSR=7.02$
	X_t-FP_{t-1}	−.23	.04	–	.73	−.40	.16	−.86	−.50	.09	$R^2=.75$
		(.13)	(.01)		(.26)	(.31)	(.33)	(.31)	(.33)	(.37)	$SSR=1.74$
Feb. 1921–	X_t	.22	–	1.15	–	–	–	–	–	–	$R^2=.04$
June 1923		(.06)		(1.10)							$D.W.=1.20$
											$SSR=2.085$
	X_t	−.56	.03	−2.20	.41	−.47	.28	−.13	.20	−.58	$R^2=.64$
		(.21)	(.01)	(1.46)	(.23)	(.26)	(.26)	(.36)	(.33)	(.31)	$SSR=.7858$
	X_t-FP_{t-1}	.22	–	–	–	–	–	–	–	–	$D.W.=1.16$
		(.05)									$SSR=2.086$
	X_t-FP_{t-1}	−.16	.04	–	.30	−.63	.44	−.52	.07	−.64	$R^2=.53$
		(.10)	(.01)		(.25)	(.27)	(.27)	(.33)	(.36)	(.34)	$SSR=.97$

Sources: The data are taken from Einzig, Statistisches Reichsamt, and several issues of the *Federal Reserve Bulletin* between 1921 and 1923. Note: SSR = sum of squared residuals.

Wallace and the author (1976) report evidence that the stochastic process which describes the evolution of the inflation rate during the German hyperinflation contained an important autoregressive component. Thus much of the variation in the one-period future inflation rate could have been predicted by forecasting on the basis of past and present inflation rates.[3] Clearly then, a potential source of additional information to use in testing the rationality of the predictions based on the forward exchange premium is the past history of the rate of inflation.

The test of the rationality of forward premium prediction will be based on estimation of the following equation:

$$(8)\quad X_t = a + b\cdot trend + \sum_{i=1}^{MLAG} c_i\cdot X_{t-i} + d\cdot FP_{t-1} + U_t$$

where $X_t \equiv$ the natural *log* approximation to the rate of inflation at time t and FP_{t-1} is defined as before. If FP_{t-1} truly provides rational expectations of X_t, then one ought certainly to accept the hypothesis that b and c_i are zero. Failure to accept the hypothesis is evidence that FP_{t-1} excludes important information available in the data.

For $MLAG=6$ the results of estimating (8) are reported in Table 1 and are evidence to reject the rationality claimed by Frenkel for FP_{t-1}. When estimated for the period from February 1921, through August 1923, the appropriate test statistic was estimated to be 5.70 which exceeds the .99 level of significance of the $F(7,22)$ distribution. In addition, inspection of the constrained regression shows that the coefficient of FP_{t-1} is estimated to be 5.21 with an estimated asymptotic standard error of 1.11 which demonstrates just how seriously the forward premium on average underpredicts the inflation rate.

One might argue, however, that the above test discards information implicit in the hypothesis that FP_{t-1} is a predictor of X_t. If the hypothesis is true, then the coefficient on FP_{t-1} in the above equations will be 1.0.

---

[3] In my dissertation, I found that I could not reject on the basis of the pairwise tests suggested by C. W. J. Granger that the rate of inflation was not Granger-caused by any in a large list of available time-series for the German hyperinflation including the forward premium.

It is clear that imposing the unit coefficient condition on the regression will raise the residual sum of squares, other things equal. But imposing the restriction may make it more difficult for other data dated t−1 and earlier to enter the regression equation. To test the rationality of FP_{t-1} as a predictor of X_t using the theoretical restriction described, an appropriate equation to estimate is

$$X_t - FP_{t-1} = a + b \cdot trend + \sum_{i=1}^{MLAG} c_i \cdot X_{t-i} + U_t \tag{9}$$

The results are reported in Table 1 and again support rejection of the hypothesis that the forward exchange premium predicted inflation rates rationally. The test statistic for the hypothesis that b and the c_i equal zero was estimated to be 9.99 which exceeds the 99 percent significance level of the $F(7,23)$ distribution.[4]

The results reported in Table 1 are evidence, moreover, to accept the hypothesis that the forward exchange premium makes no significant contribution to a prediction of X_t given past values of X. The null hypothesis is that the coefficient of FP_{t-1} in equation (4) equals zero. The test data are summarized in Table 2. At the 95 percent level of significance one would fail to reject the null hypothesis against either the alternative $d \neq 0$ or the alternative $d > 0$.

TABLE 2—TEST OF THE HYPOTHESIS THAT THE FORWARD EXCHANGE PREMIUM MAKES NO CONTRIBUTION TO A PREDICTION OF X_t GIVEN THE PAST HISTORY OF X $(MLAG=6)$

			5 Percent Critical Value	
Period	T-value	Degrees of Freedom	Two-Tail Test	One-Tail Test
Feb. 1921–Aug. 1923	1.58	22	2.074	1.717
Feb. 1921–June 1923	−1.50	20	2.086	1.725

III. Sequential Prediction of the Inflation Rate

The previous section demonstrates that by June 1923, a rational agent would not have predicted the rate of inflation solely on the basis of the forward exchange premium. This is not surprising since FP_{t-1} understates X_t in virtually every month. For example, FP_{t-1} did not exceed 1 percent per month until September 1922 even though the average monthly inflation rate for the year ending in September 1922 was 27.8 percent and the monthly rate exceeded 6.5 percent in every month but one in that year. Nevertheless, it is interesting to inquire how quickly a rational agent would have come to distrust the forward exchange premium as the sole basis for predicting the inflation rate.[5]

Two approaches were used to pursue this inquiry. First, unconstrained equations (8) and (9) and their constrained versions were estimated sequentially between February 1922 and August 1923. In each month, the regression employed only inflation rates dated earlier in order to estimate the autoregressive component of the equations and to form the statistic for the test that the data on past inflation made no significant contribution to the explanation of X_t, given FP_{t-1} and a constant.[6] The results are summarized in Table 3. For comparison, column (1) reports X_t while column (2) reports

---

[4] This result proved to be robust with respect to lag length and ending date. The reported tests were conducted varying the maximum lag of X_t in equations (8) and (9) between 1 and 6. In every case, a 95 percent test of the null hypothesis would lead to rejection. In most cases, the marginal level of significance exceeded 99 percent. The X_t series tends to explode in the second half of 1923 and thus a regression attempting to explain X_t between February 1921 and August 1923 might do well by explaining X_t for only the last few months. Thus the reported tests were conducted over the period February 1921–June 1923, a period during which I argued that the X_t series is stationary. The results for maximum lag 6 are reported in Table 1. For the test that lagged X_t and trend do not enter when the coefficient of FP_{t-1} is unconstrained, the test statistic is estimated to be 4.72 which exceeds the 99th percentile of the $F(7,20)$ distribution. If the coefficient on FP_{t-1} is constrained to 1.0, the test statistic is estimated to be 3.43 which exceeds the 95th percentile of the $F(7,21)$ distribution.

[5] The need for such a sequential approach has been discussed in Robert Barro.

[6] Due to the small number of observations available the number of lagged terms was lowered from six to four and the trend term was suppressed. In principle

TABLE 3—SEQUENTIAL PREDICTION OF X_t AND SEQUENTIAL TESTING OF THE RATIONALITY OF PREDICTIONS OF X_t BASED SOLELY ON FP_{t-1}

	X_t	FP_{t-1}	$\hat{X}_t$	X_t^*	Marginal Significance of Statistic to Test the Hypothesis $c_1=\ldots=c_4=0$ (5)	
Month	(1)	(2)	(3)	(4)	Equation (8)	Equation (9)
2/22	.113	.0003	.124	.225	.062	.335
3/22	.281	.0013	−.123	.068	.707	.713
4/22	.157	.0033	.118	.387	.661	.652
5/22	.016	.0033	.141	.467	.652	.690
6/22	.085	.0026	.147	.315	.713	.693
7/22	.358	.0026	.145	.290	.416	.419
8/22	.646	.0063	.163	.528	.326	.206
9/22	.402	.0112	.372	.932	.291	.091
10/22	.679	.0327	.230	.711	.180	.031
11/22	.712	.0667	.758	.796	.347	.004
12/22	.245	.0676	.824	.406	.321	.031
1/23	.636	.0534	.374	.154	.376	.013
2/23	.696	.0429	.640	.646	.227	.003
3/23	−.133	.2256	.427	.349	.001	.010
4/23	.064	.0330	.340	.117	.001	.010
5/23	.450	.0571	.410	.555	.001	.006
6/23	.864	.1099	−.114	.421	.015	.084
7/23	1.350	.0638	.434	1.35	.011	.012
8/23	2.536	.2852	1.239	2.03	.002	.001
9/23	3.234	.4088	3.309	1.85		
Average	.534	.056	.350	.566		
2/22–8/23	(.60)	(.077)	(.33)	(.47)		

FP_{t-1}. Column (5) reports the marginal significance levels for the F-statistics which test the hypothesis that lagged inflation rates do not help to predict the time t inflation rate given the prediction of that rate based on the forward exchange premium observed at time $t-1$. The marginal significance is the probability that the test statistic would take on a value greater than or equal to the observed value if the null hypothesis were true. Thus, in one sense, the data tell us that by no later than October 1922 an agent attempting to predict next month's inflation rate would be at least "95 percent sure" that the forward premium produced serious systematically erroneous predictions of the inflation rate.

The second method used to evaluate the quality of inflation rate predictions based solely on the forward exchange premium was to create some competitive predictions based solely on the past history of inflation. In each month the four-lag autoregression for X_t given in equation (10) was estimated by ordinary least squares and a prediction for X_{t+1} was formed under the assumption that U_t was serially uncorrelated.

$$X_t = a + \sum_{i=1}^{4} c_i X_{t-i} + U_t \tag{10}$$

These predictions, defined as the $\hat{X}_t$ series, are reported in column (3) of Table 3. In fourteen of the nineteen months between February 1922 and August 1923, $\hat{X}_t$ lies between X_t and FP_{t-1} (usually closer to X_t). Over that period X_t averaged .534, $\hat{X}_t$ averaged .350, and FP_{t-1} averaged .056. Moreover, the sum of squared prediction errors for the period was 4.95 for $\hat{X}_t$ and 17.99 for

---

for the purposes of forecasting one would wish to allow a generous number of lags and to exploit any serial correlation in the residuals. The point here is that even without taking these potentially valuable steps, it is clear that FP_{t-1} does not incorporate the information available even in the history of inflation rates.

FP_{t-1}. In short, during the period the simple autoregressive forecasting model of equation (10) decisively outperforms the forward exchange premium prediction.[7]

In summary, it is clear that a rational agent would not have predicted the inflation rate solely on the basis of the forward exchange premium. Moreover, a rational agent could be reasonably sure by November 1922 that the forward premium omitted information available in the past history of inflation to predict the future rate. Finally, such an agent would have erred less by predicting the inflation rate via a sequentially estimated autoregressive representation for X_t than by predicting on the basis of the forward premium.[8]

IV. Risk Premia and Expected Inflation

The results reported in the previous sections are disturbing because they suggest that on average it was possible for either British or German traders to exploit profitably the disparity between German and British rates. Assume again that during the hyperinflation expectations were rationally formed, that is, that π_t is equivalent to the optimal linear prediction of X_{t+1} conditioned on data available at t. The previous sections have demonstrated that

$$(11) \qquad \pi_{d,t} > FP_t$$

which implies[9] that for any positive German real rate of interest

$$(12) \qquad r_{d,t} > ln(F_t/S_t) + r_{uk,t}$$

To profit from this disparity, a British or German resident would borrow pounds in London, buy spot and sell forward marks, and lend marks in the short-term German market.

It is interesting to observe finally that weakening, in a reasonable way, the assumption of risk neutrality upon which equation (3) depends will not rehabilitate Frenkel's view that the forward premium provides expectations of inflation rational in Muth's sense. The greatest risks involved in the transactions outlined would seem to have been that German borrowers or those with short positions in forward pounds would have defaulted on their contracts. Suppose that as compensation for bearing risks, those who held long positions in forward pounds and loaned funds in German bond markets insisted that a risk premium be paid. Then equation (3) would be replaced by

$$(13) \qquad (1+r_{d,t}) = \tilde{pr} \cdot \frac{F_t}{S_t} \cdot (1+r_{uk,t})$$

where $\tilde{pr} > 1.0$ is the risk premium. By the line of reasoning employed earlier, equation (14) may be deduced as an appropriate replacement for equation (7):

$$(14) \qquad \pi_{d,t} \cong pr - \bar{r}_d + FP_t$$

where $pr = ln(\tilde{pr}) > 0$. Equation (14) says that even if agents in foreign exchange markets reacted systematically to all information useful for predicting inflation, nevertheless FP_{t-1} might on average underpredict the rate of inflation at time t.

---

[7] To make further comparison possible the predicted values from the estimation of

$$X_t = a + b \cdot trend + \sum_{i=1}^{6} c_i X_{t-i} + V_t$$

are recorded in column (4) of Table 3 as X_t^*. Thus X_t^* represents the best linear prediction possible of X_t using the information from past X when all the information in the sample is used to form estimates of a, b, and the c_is. Between February 1922 and September 1923, X_t^* averaged .566 with a sum of squared prediction errors equal to 3.55. The sum of squared errors for X_t^* was 72 percent as great as that for $\hat{X}_t$ but merely 20 percent as great as that for FP_{t-1}. Similarly, $\hat{X}_t$ dominated FP_{t-1} in the February 1922–June 1923 period discussed earlier.

[8] For corroboration, the tests of this section were all redone using the exchange data reported in Keynes and performing the appropriate interpolations for the interest rate and inflation series. With the Keynesian data, the results are even stronger to reject the rationality of an inflation prediction based solely on FP_{t-1}. The results indicate that a rational agent would not have relied on the forward premium to predict X_t from early 1922 until the end of the hyperinflation.

[9]
$$\begin{aligned} \pi_t > FP_t &= ln(F_t/S_t) + r_{uk,t} \\ \Rightarrow \pi_t &> ln(F_t/S_t) + r_{uk,t} - \bar{r}_d \\ \Rightarrow r_{d,t} &> ln(F_t/S_t) + r_{uk,t} \end{aligned}$$

But the hypothesis that agents required a risk premium to borrow in British markets and lend in German markets will not rehabilitate the view that the forward exchange premium provided rational expectations of the rate of inflation. If $\pi_{d,t-1}$ is to be the rationally formed prediction of X_t it is required that

$$X_t = \pi_{d,t-1} + W_t \tag{15}$$

where W_t is uncorrelated with any data available at t − 1. Substituting from (14) and rearranging yields

$$X_t - FP_{t-1} \cong pr - \bar{r}_d + W_t \tag{16}$$

A constant risk premium provides an explanation for the positive constants estimated for the constrained regressions of Table 1. But it cannot explain away the fact that if FP_{t-1} contained all the information available at t − 1 on the variation in X, then past X should not enter equations (8) or (9).[10]

V. Conclusions

The efficient markets hypothesis asserts that any information useful in the prediction of future spot rates will, by the market activity of agents, be built into the prices at which appropriately dated forward contracts trade. The hypothesis has two parts: the expectations hypothesis and the hypothesis that expectations are formed rationally. Thus the evidence of this paper is in fact evidence to reject the efficient market hypothesis for the market in deutsche mark exchange during the period of the German hyperinflation. A word of caution is in order, however. The analysis of this paper is based on the assumption that short-term nominal rates adjusted fully and quickly to changes in the expected rate of inflation. While this is a reasonable assumption, one would desire to test it directly by analysis of German short-term interest rates during this period. Unfortunately, none of the several statistical sources which report data for post-World War I Germany report a series of short-term market rates. In a subsequent paper I hope to readdress this topic using an index of short rates constructed from financial data available in the *Frankfurter Zeitung*.

REFERENCES

R. J. Barro, "Unanticipated Money Growth and Unemployment in the United States," *Amer. Econ. Rev.*, Mar. 1977, *67*, 101–15.

P. Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956.

Paul Einzig, *The Theory of Forward Exchange*, New York 1937.

J. Frenkel, "The Forward Exchange Rate, Expectations, and the Demand for Money: The German Hyperinflation," *Amer. Econ. Rev.*, Sept. 1977, *67*, 653–70.

C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, July 1969, *37*, 424–38.

John Maynard Keynes, "A Tract on Monetary Reform," in *The Collected Writings of John Maynard Keynes*, London 1971.

J. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961, *29*, 315–35.

M. K. Salemi, "Hyperinflation, Exchange Depreciation, and the Demand for Money in Post World War I Germany," unpublished doctoral dissertation, Univ. Minnesota 1976.

---

[10] One might argue that the appropriate risk premium did not remain constant but rather increased throughout the course of the hyperinflation. To examine this possibility constrained regressions of the form

$$X_t - FP_{t-1} = a + b \cdot trend + W_t$$

were estimated. Positive estimates of b (on the order of .02 per month) were found and are consistent with the view that the perceived riskiness of long positions in German paper assets grew as the hyperinflation progressed. Nevertheless, the hypothesis that the $c_i = 0$ in the unconstrained regression

$$X_t - FP_{t-1} = a + b \cdot trend + \sum_{i=1}^{MLAG} c_i \cdot X_{t-i} + W_t'$$

was consistently rejected at standard significance levels.

———, "Expected Exchange Depreciation and the Demand for Money in Hyperinflation Germany," *J. Money, Credit, Banking*, forthcoming.

——— **and T. J. Sargent,** "The Demand for Money During Hyperinflation Under Rational Expectations: II," *Int. Econ. Rev.*, Oct. 1979, *20*, 741–58.

T. J. Sargent and N. Wallace, "Rational Expectations and the Dynamics of Hyperinflation," *Int. Econ. Rev.*, June 1973, *14*, 328–50.

Board of Governors of the Federal Reserve System, *Fed. Reserve Bull.*, Washington, various issues.

Statistisches Reichsamt, *Zahlen zur Geldenwertung in Deutschland 1914 bis 1923*, Berlin 1925.

The Forward Exchange Rate, Expectations, and the Demand for Money— The German Hyperinflation: Reply

By JACOB A. FRENKEL*

Standard theories of the demand for money assert that the demand depends on the expected cost of holding money. To the extent that money is held as a substitute for securities, the relevant cost is the rate of interest; to the extent that money is held as a substitute for ownership of commodities, the relevant cost is the expected rate of inflation; and to the extent that (domestic) money is held as a substitute for foreign exchange, the relevant cost is the expected change in the exchange rate. In my earlier 1977 paper dealing with the demand for money during the German hyperinflation of 1921–23, I used the forward premium on foreign exchange as a direct measure of the expected cost of holding German marks. This approach reflects the belief that in a highly inflationary economy with flexible exchange rates one of the important substitutes for the holdings of domestic money is foreign exchange. Indeed, the evidence from the German hyperinflation illustrated the substitutability between foreign exchange and domestic money. The accelerated inflation resulted in a gradual replacement of domestic money by foreign exchange in performing the traditional roles of money as a unit of account, as a store of value, and even as a medium of exchange. The major virtue of the proposed direct measure of expectations is that it is not derived from a specific mechanistic formula, but rather, it reflects the (unobservable) expectations of agents as manifested in market prices. Prior to incorporating this measure of expectations in estimating the demand for money, I explored some of the efficiency properties of the market for foreign exchange by examining the relationship between spot and forward exchange rates. I concluded that even though the forward rate seems to have underpredicted the actual course of events, the overall relationship between spot and forward rates lent support to the approach of using data from that market as the basis for inference on expectations.

In his comment, Michael Salemi argues that the foreign exchange market did not function efficiently since the forward premium on foreign exchange did not contain all the information available to agents to predict the rate of inflation. In Section I, I show that using Salemi's own techniques for testing efficiency one *cannot* reject the hypothesis that the foreign exchange market functioned efficiently. In particular I show that one cannot reject the hypothesis that the forward premium on foreign exchange contained all the information available for predicting the rate of depreciation of the mark in terms of foreign exchange. Section II contains some further analysis of the relationship between spot and forward exchange rates with concluding remarks in Section III.

I. The Rationality of Using the Forward Premium as a Measure of Expectations

The fundamental difference between Salemi's framework and mine lies in the interpretation of the margin of substitution that is measured by the forward premium on foreign exchange. In my analysis the forward premium measures the expected depreciation of the currency in terms of foreign exchange, while in Salemi's analysis it measures the depreciation in terms of goods (or more precisely it measures the difference between German and *U.K.* inflation rates). As is clear in the long run, when various parity conditions are satisfied, the two measures coincide. Salemi's procedure assumes that all parity conditions are satisfied. Thus he assumes that (i) the interest parity theory holds, that is, the difference between German and British rates of interest equals the forward premium on foreign exchange; (ii) Irving Fisher's relationships hold, that is, the nominal rates of interest in Germany and in the United Kingdom equal the sum of the respective real rates of interest and the expected rates of inflation; (iii) the real rates of interest in Germany and Britain are constant; (iv) the real rates of interest in Germany and Britain are equal to each other; and (v) that relative purchasing power parity holds so that the percentage change in the exchange rate equals the difference between expected rates of inflation in Germany and Britain. Of course, not all of the above conditions are independent. If some of these conditions are not satisfied in the short run, the forward premium on foreign exchange may not be used to measure the expected (differential) rate of inflation even though it may be used as a measure of the expected depreciation of the currency in terms of foreign exchange.

---

*University of Chicago and National Bureau of Economic Research, Inc. I am indebted to Craig S. Hakkio for comments and research assistance, and to financial support from a National Science Foundation grant, SOC 78-14480. The research here is part of NBER's research program in International Studies. Any opinions expressed are my own and not those of the NBER.

The results of Salemi's test confirm the hypothesis that the forward premium on foreign exchange is not a sufficient statistic for the prediction of future *inflation* since the past history of the rate of inflation "helps" the prediction. This, however, is *not* the test that is relevant for my conceptual framework. For that framework, the relevant test would be whether the forward premium on foreign exchange is a sufficient statistic for the prediction of future *depreciation* of the currency in terms of foreign exchange in the sense that, given the value of the forward premium, the past history of the rate of depreciation does not help the prediction.[1] Applying Salemi's test procedure to this question requires the estimation of the following equation:

$$(1)\quad X_t=\alpha_0+\alpha_1 t+\sum_{i=1}^{M}\beta_i X_{t-i}+\gamma\pi^*_{t-1}+u_t$$

where X_t denotes the percentage change in the spot exchange rate $(\ln S_t-\ln S_{t-1})$, π^*_{t-1} denotes the forward premium on foreign exchange $(\ln F_{t-1}-\ln S_{t-1})$, t denotes time, S and F denote the spot and the forward exchange rates, respectively, M denotes the number of lags, and u denotes an error term. If π^*_{t-1} is a rational expectations measure of the percentage depreciation of the currency X_t, then one should not be able to reject the hypothesis that α_1 and β_i are zero. The results of these tests for various numbers of lags are reported in Table 1; also reported in Table 1 are the results of testing the joint hypotheses that α_1 and β_i are zero and that γ, the coefficient of the forward premium, is unity. The relevant statistic for testing the null hypothesis is an *F*-statistic which is reported in Table 1. As is evident in all cases the null hypothesis cannot be rejected at the 95 percent confidence level since the values of the various *F*-statistics fall well below the corresponding critical values. It is concluded therefore that the forward premium on foreign exchange *may*

---

[1] In addition to this fundamental conceptual difference in the interpretation of the role of the forward premium, it is noteworthy that Salemi's measure of what he terms "the Frenkel inflation predictor" (defined as the forward premium plus the foreign nominal rate of interest minus the domestic real rate of interest) was never defined or used by me. Furthermore, Salemi's derivations assume that $\ln(1+x)$ can be approximated by x. For the period of the hyperinflation, in particular during 1923, this approximation is extremely poor since the value of x (the sum of the real rate of interest and the expected rate of inflation) is relatively large. For example, the regression of the actual wholesale price index rate of inflation DP/P on $\ln(1+DP/P)$ yields

$$DP/P=\underset{(.029)}{-.073}+\underset{(.075)}{1.689}\ln(1+DP/P)$$

and thus indicates the extent of the error that is introduced by the approximation. As an historical aside it should be noted that the name of the German currency during that period was the "mark" and *not* the "deutsche mark" as used by Salemi. The deutsche mark was introduced much later when it replaced the reichsmark.

TABLE 1—TEST OF RATIONALITY OF FORWARD PREMIUM PREDICTION OF CURRENCY DEPRECIATION (Monthly Data, February 1921–August 1923)

Null Hypothesis	Number of Lags	F-statistic
$\alpha_1=0, \beta_i=0$	4	$F(5,18)=.978$
	5	$F(6,17)=1.209$
	6	$F(7,16)=1.206$
$\alpha_1=0, \beta_i=0,$ $\gamma=0$	4	$F(6,18)=1.442$
	5	$F(7,17)=1.552$
	6	$F(8,16)=1.653$

Note: Data are from Paul Einzig.

be viewed as a rational expectations measure of the percentage depreciation of the currency. These results indicate that, to the extent that foreign currency holdings are viewed as a relevant substitute for the holdings of domestic money, the forward premium on foreign exchange may be viewed as a proper rational expectations measure of this relevant margin of substitution.[2]

II. Further Evidence on the Relationship between Spot and Forward Rates

In my 1976, 1977, and 1979 papers, I analyzed the efficiency of the foreign exchange market during the German hyperinflation by examining the relationship (obtained by ordinary least squares (*OLS*) estimation) between the (logarithms of the) current spot exchange rate and the one-month forward exchange rate prevailing at the previous month as in equation (2):

$$\ln S_t = a + b \ln F_{t-1} + u_t \qquad (2)$$

If the forward exchange rate is an unbiased forecast of the future spot exchange rate, then the constant should not differ significantly from zero and the slope coefficient should not differ significantly from unity.[3] Efficiency requires that the residuals from the estimated regression contain no relevant information and thus, that they are serially uncorrelated. My inference on efficiency was based on the analysis of equation (2) which stimulated Salemi's comment. In this section I deal with two technical issues associated with this procedure.

A. Errors in Variables

One of the assumptions underlying the use of equation (2) is the notion that the forward exchange rate may serve as a good proxy for the unobservable value of the expected future spot exchange rate. This assumption provided the justification for using equation (2) instead of the more fundamental relationship that is embodied in equation (3):

$$\ln S_t = E_{t-1} \ln S_t + \varepsilon_t \qquad (3)$$

where $E_{t-1} \ln S_t$ denotes the expected spot exchange rate for period t as of period t − 1. If, however, the forward exchange rate at t − 1 is a "noisy" proxy for the expected future value of the spot rate, we would obtain that

$$a + b \ln F_{t-1} = E_{t-1} \ln S_t + v_{t-1}; \; E(v_t) = 0 \qquad (4)$$

---

[2] The fact that the forward premium may be viewed as a proper measure of the expected depreciation of the mark in terms of foreign exchange but not as a measure of the expected domestic (wholesale price) inflation stems from short-run deviations from purchasing power parities. The existence of these deviations led Andrew Abel et al. to estimate the demand for money as a function of both margins of substitution—the expected change in the exchange rate and the expected rate of inflation. They showed that the coefficient of the external margin of substitution was about three times the size of the coefficient of the internal margin of substitution. In his conclusion Salemi emphasizes that his analysis is based on the assumption that "short-term nominal rates [of interest] adjusted fully and quickly to changes in the expected rate of inflation" and that there is a difficulty in testing this since "none of the several statistical sources which report data for post-World War I Germany report a series of short-term market rates" (p. 769). As a matter of fact James Angell (pp. 370–71) reports two series of money rates (private discount up to December 1922 and day-to-day rates for the entire period). These rates are low relative to the rate of inflation and thus shed some doubts on the maintained hypothesis concerning the efficiency of the German credit market during the hyperinflation.

[3] More precisely, the requirement that the mathematical expectations of S_t and F_{t-1} be equal (rather than the requirement of equality between the mathematical expectations of the logarithms of S_t and F_{t-1}) imply that the constant term in equation (2) should not differ significantly from $-0.5\sigma_u^2$.

TABLE 2—TEST FOR THE PRESENCE OF ERROR IN VARIABLES AND EFFICIENCY OF THE FOREIGN EXCHANGE MARKET (Monthly data, February 1921–August 1923)

Method of Estimation	Dependent Variable	Constant	lnF_{t-1}	R^2	*s.e.*	*D.W.*	*m*
OLS	lnS_t	−0.454	1.095	.98	.456	1.89	
		(0.254)	(0.029)				
IV	lnS_t	−0.401	1.088	.98	.457	1.88	.38
		(0.268)	(0.030)				

Note: Standard errors are in parentheses below each coefficient; *s.e.* is the standard error of the equation. The *IV* estimation method is used in order to allow for the possibility of errors in variables arising from using $ln\,F_{t-1}$ as a proxy for the expected future spot rate; the instruments are lagged values of the dependent and independent variables, a constant, time, and time squared. The *m*-statistic tests for the absence of errors in variables. The *m*-statistic is distributed χ^2 with two degrees of freedom. The critical value for $\chi^2(2)$ is 5.99 (95 percent) and 9.21 (99 percent).

and substituting equation (4) into equation (3) yields

$$(5) \qquad ln\,S_t = a + b\,ln\,F_{t-1} + (\varepsilon_t - v_{t-1})$$

In this case the error term in equation (2) would be $u_t = \varepsilon_t - v_{t-1}$, the assumption that $cov(u_t, ln\,F_{t-1}) = 0$ would entail a specification error, and the application of the *OLS* procedure would yield biased estimates due to the classical errors in variables bias.

In order to examine the possibility that the *OLS* estimates might be subject to the errors in variables bias, I test in what follows the hypothesis that $cov(u_t, ln\,F_{t-1}) = 0$. This test follows the specification test outlined by Jerry A. Hausman; it was recently applied by Maurice Obstfeld to the analysis of the foreign exchange market during the 1970's. To perform the test, equation (2) was estimated by applying the *OLS* procedure as well as by using an instrumental variables (*IV*) estimation method. Under the null hypothesis of no misspecification, the *OLS* coefficients vector $\hat{b}_0$ is an efficient and an unbiased estimate of the true coefficient vector. Under the alternative hypothesis of misspecification, the vector $\hat{b}_0$ would be biased and an unbiased coefficient vector $\hat{b}_1$ could be obtained by applying an instrumental variables estimation procedure. The test statistic relevant for testing the null hypothesis can be written as

(6)

$$m = (\hat{b}_1 - \hat{b}_0)'(var\,\hat{b}_1 - var\,\hat{b}_0)^{-1}(\hat{b}_1 - \hat{b}_0)$$

where *var* $(\hat{b}_1)$ and *var* $(\hat{b}_0)$ denote the variance-covariance matrices of $\hat{b}_1$ and $\hat{b}_0$, respectively. Under the null hypothesis, *m* is distributed (in large samples) as χ^2 with two degrees of freedom. Table 2 reports the results of estimating equation (2) by applying *OLS* and *IV* estimation methods. As may be seen the two vectors of coefficients $\hat{b}_1$ and $\hat{b}_0$ are very close to each other (the constants are −.454 and −.401 and the slopes are 1.095 and 1.088). Consequently, the resulting *m*-statistic is .38 which is well below 5.99, the critical value of $\chi^2(2)$ at the 95 percent confidence level. It is concluded, therefore, that the use of the forward exchange rate as a proxy for expectations does not introduce a significant errors in variables bias and thus the use of the *OLS* estimation procedure seems appropriate.

B. The Assumption of a Normal Distribution

One of the key assumptions underlying the statistical tests and the resulting inference is the assumption that the distribution generating the sample is normal. In a recent paper dealing with the hyperinflation William Krasker showed that if participants in the foreign exchange market believed that there was some probability that at some point monetary reform would stop the hyperinflation, then the normality assumption underlying equation (2) would not be satisfied since the residuals would not be serially independent. Under these circumstances the standard *t*-tests for significance of coefficients would not be appropriate. In what follows I apply various tests of the normality assumption to the case at hand.

To test the hypothesis that the distribution generating the sample is normal, I computed the studentized range which measures the ratio of the range of the residuals to their standard error. The resulting statistic was 4.61 which lies within the interval 3.37–5.06, which is the relevant range for the 95 percent confidence level. Thus, based on this test, the normality assumption cannot be rejected. In addition, the value of the skewness of the distribution (the ratio of the third moment around the mean to the third power of the standard deviation), which assumes the value of zero for the normal case, is $-.431$ with a standard error of .421. Likewise, the excess value of the kurtosis of the distribution (the difference between 3 and the ratio of the fourth moment around the mean to the fourth power of the standard deviation), which assumes the value of zero for the normal case, is .957 with a standard error of .821.[4] Thus, both the skewness and the kurtosis measures are consistent with the hypothesis that the distribution generating the sample is normal.

III. Concluding Remarks

I have shown that the forward premium on foreign exchange during the German hyperinflation may be viewed as a rational predictor of the *depreciation* of the currency in the sense that, given the value of the forward premium, the past history of the rate of depreciation does not help the prediction. Salemi's result that the forward premium may not be viewed as a rational predictor of *inflation* suggests that in the short run, when various parity conditions may not hold, the expectations that are reflected in the forward premium on foreign exchange need to be interpreted with care. In particular, the forward premium reflects expectations concerning the depreciation of the currency in terms of foreign exchange rather than in terms of domestic goods. In the context of the demand for money, the results indicate that the expected holding cost depends on the specification of the relevant substitutes for the holdings of domestic money, and that the forward premium should be viewed as a measure of the external (rather than the internal) margin of substitution. It is important to emphasize that short-run deviations from purchasing power parities need not indicate inefficiency of markets. Consequently, tests which require that all parity conditions hold (like tests which examine the performance of the forward premium in predicting domestic inflation) do not seem to be appropriate tests of efficiency for the short run.

REFERENCES

A. Abel et al., "Money Demand During Hyperinflation," *J. Monet. Econ.*, Jan. 1979, *5*, 97–104.

James W. Angell, *The Recovery of Germany*, New York 1929.

Paul Einzig, *The Theory of Forward Exchange*, London 1937.

R. A. Fisher, *Statistical Methods for Research Workers*, 14th ed., Darien 1970.

J. A. Frenkel, "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence," *Scand. J. Econ.*, No. 2, 1976, *78*, 200–24.

______, "The Forward Exchange Rate, Expectations, and the Demand for Money: The German Hyperinflation," *Amer. Econ. Rev.*, Sept. 1977, *67*, 653–70.

______, "Further Evidence on Expectations and the Demand for Money During the German Hyperinflation," *J. Monet. Econ.*, Jan. 1979, *5*, 81–96.

J. A. Hausman, "Specification Tests in Econometrics," *Econometrica*, Nov. 1978, *46*, 1251–71.

W. A. Krasker, "The 'Peso Problem' in Testing the Efficiency of Forward Exchange Markets," *J. Monet. Econ.* Apr. 1980, *6*, 269–76..

M. Obstfeld, "Expectations and Efficiency in the Foreign Exchange Market," unpublished paper, M.I.T. 1978.

M. K. Salemi, "The Forward Exchange Rate, Expectations, and the Demand for Money—The German Hyperinflation: Comment," *Amer. Econ. Rev.*, Sept. 1980, *70*, 763–70.

---

[4] The standard errors of the estimated skewness and kurtosis are computed using the formulae in R. A. Fisher, p. 75.

Exchange Rate Dynamics, Portfolio Balance, and Relative Prices

By ROBERT DRISKILL*

Much of the recent literature on exchange rate determination has emphasized the role of the exchange rate in maintaining continuous portfolio balance among existing stocks of financial assets. This asset equilibrium approach has been referred to as the "modern" or "new" view because it assigns the exchange rate no role in balancing the flow demands and supplies of foreign exchange arising from trade in goods and capital, the point of departure in traditional exchange rate theory.

This paper attempts to extend the asset equilibrium approach by incorporating in it this neglected, more traditional aspect of exchange rate determination. While this extension leaves unaltered the basic asset equilibrium conclusions about full stock/flow equilibrium relationships between nominal exogenous asset supplies, exchange rates, and price levels, it does substantially change the conclusions about exchange rate and price level behavior in the adjustment phase between full stock/flow equilibria. In particular, the short-run response of the exchange rate to a step change in the money supply need not be an "overshoot" of its long-run value, as it must be in the models of Pentti Kouri and Rudiger Dornbusch. Rather it may be a much smaller response if certain plausible parameter values obtain. In addition, the intermediate-run dynamic price and exchange rate adjustments to a money supply step change are also shown to be substantially different in character than the monotonic adjustment paths implied by the asset equilibrium theories.

This analysis is not the first to incorporate trade flow responses to relative price changes into a model of exchange rate determination. Guillermo Calvo and Carlos Rodriguez analyze a portfolio-balance model with traded and nontraded goods. Their model, however, predicts instantaneous adjustments to full stock/flow equilibrium in response to step changes in the money supply; the only exchange rate dynamics in their model occur because of changes in rates of growth of money supplies. Apart from the question of whether such instantaneous price level adjustments are realistic, this characteristic makes their model not suitable for comparison with the important asset equilibrium model of Dornbusch.

A second model which incorporates trade flows into exchange rate determination is that of Jürg Niehans (1977). His model, though, departs from the asset equilibrium approach in that individuals do not continuously maintain portfolio balance among existing assets, but rather adjust each asset slowly to its desired level. While the implications of his model for the intermediate-run dynamics and for full stock/flow equilibria are similar to those found in the present analysis, the short-run implications are different because of the assumption of lagged adjustment for each individual asset.

The present analysis, then, will focus on the joint adjustment of the exchange rate and the domestic price level to a step change in the domestic money supply under conditions of continuous portfolio balance. The model used resembles that of Dornbusch in that the domestic price level is "sluggish," that is, it takes time for the price level to adjust from one full stock/flow equilibrium to another. It also follows Dornbusch in that the domestic-foreign interest differential is assumed equal to the expected instantaneous or one-period exchange rate change, that is, there is perfect capital mobility. It departs from the Dornbusch model, however, by explicitly incorporating wealth into the money demand function. It also departs from his model, and in turn resembles Niehans' model, in that the slow adjustment of the domestic price level to full equilibrium is the result of individuals' attempts to slowly adjust their stock of wealth to its desired level. This eliminates many of the well-known problems involved in analyzing flow market disequilibrium, problems which are inherent in the Dornbusch price equation formulation of prices responding slowly to excess demand in the flow market for goods. In the present analysis, all prices are market clearing at every period of time. The emphasis is on *stock* rather than *flow* disequilibrium.[1]

---

*Assistant professor, University of California-Davis. I would like to thank Jürg Niehans, Bela Balassa, and George Borts for helpful comments on an earlier draft.

I. The Model

A. *General Features*

In order to accentuate the critical roles of relative prices and portfolio balance, the analysis abstracts from many presumably important aspects of economic reality. A strong attempt is made, however, to focus on those results which are probably valid even in the context of a more general model.

The model is of a small economy, facing a fixed foreign price level set for simplicity at one, and a fixed interest rate on foreign bonds r^*. Domestic output, denoted by y, is assumed fixed, and an imperfect substitute for goods produced in the rest of the world. There are no physical assets, and domestic residents hold only two financial assets: domestic money M, exogenously determined by the government, and foreign bonds. The foreign currency value of domestically held foreign bonds will be denoted by F. Foreigners are assumed to hold no M.

There are four basic building blocks in the model: a portfolio balance condition describing how individuals allocate their existing stock of wealth between money and bonds; a specification of a stock demand for real wealth; a specification of how individuals adjust actual real wealth to its desired level; and a foreign exchange market equilibrium condition. For purposes of analyzing the long-run effects of monetary changes, the model is specified in a general functional form. To analyze the dynamic adjustment from one full equilibrium to another, a specific functional form is specified which permits a complete analytic solution. The following three subsections develop the four building blocks in detail.

B. *Portfolio Balance*

To facilitate development of the main points of the analysis, the asset demand equations are assumed: (a) to have zero income elasticities, (b) to be homogenous of degree one in wealth, and (c) to depend on the nominal domestic rate of interest r. Assuming asset market equilibrium obtains every period, the market-clearing expression can be written as

$$(1)\qquad M/EF=Q(r),\ Q'<0$$

where E is the domestic currency price of foreign exchange. For simplicity, $Q(r)$ is assumed to have the following functional form:

$$(2)\qquad Q=C\exp(-r)$$

where C is a constant. Note that the assumption of degree one wealth homogeneity is rather restrictive. Consequently, the implications of relaxing this assumption will be discussed at the appropriate point in the analysis.

Since there are no domestic bonds, the "domestic" interest rate r is simply the definition of the return from holding foreign assets, and is equal to r^*+x, where x is the expected percentage rate of change of E. The variable x is assumed to be a linear function of the gap between $\bar{e}$, the *log* of the long-run equilibrium exchange rate, and e, the *log* of the current rate:

$$(3)\qquad x=\theta(\bar{e}-e)$$

---

[1] For another analysis which focuses on stock rather than flow disequilibrium as a more fruitful motivation for sluggish price behavior, see Douglas Purvis.

This expectational scheme, which is identical to that used by Dornbusch, simply means that the exchange rate is expected to monotonically return from its current value to its long-run equilibrium value at a rate proportional to the gap between the current and long-run rate.[2] Note that $\bar{e}$ is assumed known and is proportional to the current money supply.

C. *Wealth*

For simplicity, desired wealth is assumed to be a linear function of the (constant) value of permanent real income.

$$W^*/P=K \tag{4}$$

where K is a constant and P is the domestic currency price of domestic output.[3]

The speed with which people adjust actual real wealth to its desired level is a positive function of the gap between desired and actual real wealth, and is assumed to be *log*-linear in form:

$$Dw-Dp=\lambda[w^*-p-(w-p)] \tag{5}$$

where $w=\log W$ and $p=\log P$. This equation relates the rate of change of the price level to rates of change of the domestic currency value of financial asset stocks and to the magnitude of stock disequilibrium in the economy. It forms the basis of the price adjustment equation developed later. The underlying assumption here is that there are adjustment costs in quickly spending accumulated financial wealth on goods. From a fundamental point of view these adjustment costs should be related to search costs and uncertainty, a task beyond the scope of this paper. Note that, in this type of model, real expenditure, or real income minus real savings, is not a decision variable. The decision variable is the speed of adjustment of actual real wealth to its desired level.[4]

Finally, actual wealth will be approximated around equilibrium by a Cobb-Douglas function:[5]

$$W\approx AM^{\alpha}(EF)^{1-\alpha} \qquad 0<\alpha<1 \tag{6}$$

D. *Foreign Exchange Market Equilibrium*

To complete the model, equilibrium conditions in the foreign exchange market must be specified. Under floating rates, the trade balance plus interest payments from abroad must equal the change in private foreign assets. The trade balance, measured in units of foreign currency, is similar to that used by Niehans, which, when expressed as a percentage of foreign assets, is assumed to be linear in the *log* of relative prices:

$$T/F=\beta(e-p)+b \tag{7}$$

$\beta>0$, b a constant

Since $DF=T+r^*F$, this implies that

$$Df=\frac{d\log F}{dt}=\beta(e-p)+b' \tag{8}$$

b' a constant

This specification implicitly assumes that the income effects on the trade balance are so small they can be ignored. This seems not too high a price to pay for the resulting simplifications. Furthermore, it does not affect the basic conclusion of the dynamic analysis, that even a simple model predicts complex adjustment paths. Note that the trade balance is assumed to always satisfy the Marshall-Lerner conditions ($T_1 > 0$). Incorporating a more realistic lagged response of traded goods to relative price changes would have created a much more complex model and obscured the other points of the analysis.

---

[2] Note that, unlike the Dornbusch model, equation (3) in this model can never be consistent with perfect foresight. While it would be interesting to investigate the model's behavior under perfect foresight, this would add an extra dimension to the dynamic system developed in Section IV, making it analytically intractable.

[3] Desired real wealth could be specified as a decreasing function of the expected rate of inflation, or could be deflated by an index of P and E rather than only P, without changing any of the qualitative conclusions of the analysis.

[4] For a discussion of the relation ship of this type of model to the traditional consumption or expenditure function, see Niehans (1978, p. 236).

[5] To make W a good approximation to actual wealth $\overline{W}$, α must be chosen so that $\alpha/(1-\alpha)=\overline{Q}$, where $\overline{Q}$ is the equilibrium ratio of money to bonds M/EF, and A must be chosen to equal $\alpha^{-1}\overline{Q}^{1-\alpha}$. To see this, consider a linear expansion of W around (M_0, B_0), where $M_0=\overline{Q}B_0$: $\overline{W}\approx[\alpha A(\overline{Q})^{\alpha-1}]M+[1-\alpha)A(Q)^{\alpha}]B$. For W to equal $\overline{W}$, α and A must be chosen so that: (i) $A(Q)^{\alpha-1}=1$; (ii) $(1-\alpha)A(Q)^{\alpha}=1$. Hence, $\alpha/(1-\alpha)=\overline{Q}$, and $A=\alpha^{-1}\overline{Q}^{1-\alpha}$.

Equations (1)–(8) determine the values of E, P, and F, and also the time paths dP, dE, and dF, given M and initial conditions on E, P, and F. The remainder of the paper discusses this determination in terms of full stock/flow equilibrium, and in terms of the adjustment from one full equilibrium to another. To avoid using Laplace transforms, which are necessary to solve continuous-time differential equations subject to instantaneous shocks such as a timeless step change in the money supply, the analysis of the adjustment phase of the economy will be carried out using discrete time.

II. Full Equilibrium

In full equilibrium, all stocks are at their desired levels, and, in the present context, the price level and the exchange rate are constant. Full equilibrium can then be described by the following equations;[6]

$$\frac{M}{P} = L(r^*) \qquad L_1 < 0 \tag{9}$$

$$\frac{EF}{P} = G(r^*) \qquad G_1 > 0 \tag{10}$$

$$\frac{T}{F}\left(\frac{E}{P}\right) + r^* = 0 \qquad T_1 > 0 \tag{11}$$

The comparative statics of this system of equations are well known and easily understood: money supply changes lead to equiproportional changes in the exchange rate and the price level, but leave F unchanged. That is, the quantity theory and purchasing power parity hold in the long run.[7] In terms of *log P* and *log E*, denoted by p and e, respectively, the pairs of (e, p) corresponding to full equilibrium must lie on the straight line A in Figure 1. If, starting from full equilibrium at point $(\hat{e}, \hat{p})$, the money supply is increased by Δm, the new full equilibrium will be at $(\bar{e}, \bar{p})$. Line A may be thought of as the trade equilibrium or purchasing power parity line.

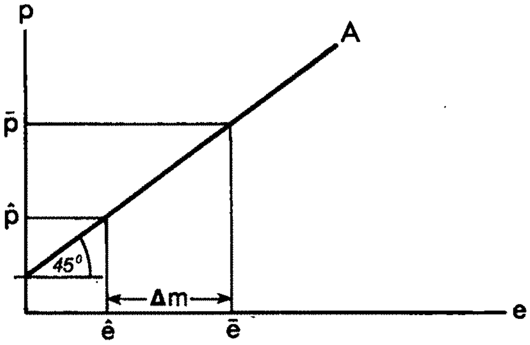


FIGURE 1. THE TRADE EQUILIBRIUM OR PURCHASING POWER PARITY LINE

III. Adjustments to Full Equilibrium: Initial Effects

The question, in terms of Figure 1, is how p and e adjust from point $(\hat{e}, \hat{p})$ to point $(\bar{e}, \bar{p})$. The interesting issues are what determines the magnitude of relative price deviations from the purchasing power parity line, and what determines whether the adjustment path is monotonic or cyclical in terms of the domestic price level, the exchange rate, or both. In order to highlight results concerning overshooting and to compare these results with those of the existing asset equilibrium models, the analysis of this adjustment is broken down into two parts:

---

[6] Note that wealth does not appear in the long-run asset demand functions; since it is at its desired level, it can be "solved out." See G. C. Archibald and Robert Lipsey, Kouri, or Niehans (1978) for a fuller discussion. Note also that (9) and (10) are not the same as (1) and (2).

[7] Note that the existence of government debt would require a revision of this homogeneity principle: equiproportional changes in money *and* government bonds would lead to equiproportional changes in the exchange rate and the price level.

1) initial period effects and 2) the following intermediate-run dynamics to full stock/flow equilibrium. This section discusses the initial period effects. Note that by using discrete-time rather than continuous-time analysis, a methodological problem encountered by Niehans in analyzing the initial period movement in prices and exchange rates is sidestepped. In Niehans' continuous-time analysis, it was only possible to show that the exchange rate and price level must, following a step change in the money supply, instantaneously jump to somewhere along a specified line in (e, p) space; the actual values of e and p could only be guessed by using as clues plausible economic considerations not part of the formal model. Here, the actual initial period values of e and p are well-specified and easily computed.

To determine the initial response of the model to a step change in the money supply at, say, $t=0$, first consider the full stock/flow equilibrium conditions holding at $t=-1$. For simplicity, and without loss of generality, *log K* from equation (4) and b' from equation (8) are set equal to zero. At $t=-1$, the definition of wealth is

$$w_{-1}=\alpha m_{-1}+(1-\alpha)(e_{-1}+f_{-1}) \tag{12}$$

where $m=\log M$. The equality of desired and actual wealth at $t=-1$ implies

$$w_{-1}=p_{-1} \tag{13}$$

Since at $t=-1$ the economy is assumed in full stock equilibrium, f is unchanging, and foreign payments equilibrium implies

$$e_{-1}=p_{-1} \tag{14}$$

Now consider an unexpected overnight increase in the money supply from level $\hat{m}$ at $t=-1$ to $\bar{m}$ at $t=0$. Taking the logarithm of equation (1), note that portfolio balance at $t=0$ requires

$$\bar{m}-e_0-f_0=c-r^*-\theta(\bar{e}_0-e_0), \tag{15}$$

where $c=\log C$ and $\bar{e}$ is the new long-run equilibrium value of e based on the information that $m=\bar{m}$. Without loss of generality, $c-r^*$ is assumed equal to zero. Noting that $\bar{m}=\bar{p}=\bar{e}$, where $\bar{p}$ and $\bar{e}$ are the new long-run equilibrium levels of p and e, respectively, and noting that the stock-adjustment equation for wealth, equation (5), should be interpreted as

$$w_0-\hat{w}=p_0-p_{-1}+\lambda\left[(w_t^*-p_t)-(\hat{w}-p_{t-1})\right] \tag{16}$$

where $\hat{w}=\alpha\bar{m}+(1-\alpha)(e_{-1}+f_{-1})$, the values of e_0 and p_0 can be computed as[8]

$$e_0=\bar{m}-\frac{\beta\alpha(1-\lambda)}{1+\theta+\beta+\beta\theta(1-\alpha)}(\bar{m}-\hat{m}) \tag{17}$$

$$p_0=\bar{m}-\frac{(1+\theta+\beta)\alpha(1-\lambda)}{1+\theta+\beta+(1-\alpha)\beta\theta}(\bar{m}-\hat{m}) \tag{18}$$

Since α, λ, and θ are fractions, the monetary expansion of size $(\bar{m}-\hat{m})$ creates a less than proportional depreciation of the exchange rate, that is, the exchange rate undershoots its long-run equilibrium value. This contrasts with the overshooting of Dornbusch. The critical difference between the two models which is responsible for these different results is the exclusion of wealth as an argument in Dornbusch's money demand equation. To see this, consider the reason for overshooting in the Dornbusch model. There, the higher money demand necessary to match an increased money supply only comes about via a lower domestic interest rate. Given that the domestic interest rate equals the foreign rate plus the expected percentage exchange rate change, there must be an expected exchange rate appreciation. With the expectational assumption of equation (3), this means the current exchange rate must depreciate beyond its long-run equilibrium value.

---

[8] The idea that w is computed using the new level of the money supply expresses the idea that, at the beginning of the period, a nominal increase in money balances is reflected as an increase in real balances. It is this observed beginning of period discrepancy between actual and desired real balances that causes prices to react as people partially adjust to full equilibrium.

Once wealth is an argument in the money demand function, though, any increase in the exchange rate increases the value of domestically held foreign assets and thus increases money demand.[9] Thus a monetary expansion leads to a capital outflow as people readjust portfolios. This outflow in turn depreciates the exchange rate, automatically raising money demand and choking off excess demand in the market. Indeed, in the case analyzed here where the demand for money is homogeneous of degree one in wealth, this "wealth effect" is so strong as to create excess supply in the money market, requiring a rise in the domestic interest rate or, equivalently, an expected exchange rate depreciation.[10]

Another aspect of the initial period effect is that (e_0-p_0) is positive, implying that a monetary expansion leads initially to a trade surplus. Furthermore, differentiation of (17) and (18) with respect to β shows that $de_0/d\beta>0$ and $dp_0/d\beta>0$. This means that the larger is the trade balance response to relative price changes, the closer e_0 and p_0 lie to the purchasing power parity line. In the limit as $\beta \to \infty$, e_0 and p_0 both approach the value

$$\bar{m}+\Delta m\left\{\frac{-\alpha(1-\lambda)}{1+(1-\alpha)\theta}\right\}$$

so that the economy at $t=0$ remains at purchasing power parity, albeit at higher levels of e and p. Intuitively, as β becomes bigger, foreign exchange transactions are dominated by trade account considerations, and the market-clearing exchange rate must approximate that which equilibrates the trade balance alone. This is a theme which reappears in the analysis of intermediate-run dynamics.

Finally, note that as λ, which measures the speed of real wealth adjustment, becomes larger, e_0 and p_0 move closer to $\bar{e}$ and $\bar{p}$, in fact reaching $\bar{e}$ and $\bar{p}$ when $\lambda=1$.

IV. Adjustments to Full Equilibrium: Intermediate-Run Dynamics

This section analyzes the dynamic adjustment of e and p to full stock/flow equilibrium following the initial adjustment to (e_0, p_0). The concern is how simple or complex is this adjustment phase, and what determines whether deviations from purchasing power parity are large or small.

To pursue the analysis, difference equations in e and p are derived as follows. First, by taking the first difference of the logarithm of (1), substituting $\beta(e_t-p_t)$ for Δf_t, and noting that, for $t>0$ Δm and $\Delta\bar{e}$ are zero, the following difference equation is formed:

$$(19)\quad e_t-e_{t-1}=\frac{-\beta}{1+\theta}e_t+\frac{\beta}{1+\theta}p_t$$

Second, using the wealth-adjustment equation (5), and again substituting $\beta(e_t-p_t)$ for Δf_t and noting that $\Delta m=\Delta\bar{e}$ for $t>0$, the following difference equation is also formed:

(20)

$$p_t-p_{t-1}=\frac{-\lambda(1+\theta+\beta)+\beta\theta(1-\alpha)}{1+\theta+\beta+\beta\theta(1-\alpha)}p_{t-1}+\frac{-\lambda(1-\alpha)\theta(1+\theta+\beta)+\beta\theta(1-\alpha)}{1+\theta+\beta+\beta\theta(1-\alpha)}e_{t-1}+N$$

where N is a constant. Equations (19) and (20) constitute a coupled system of first-order difference equations which can be shown to be everywhere stable. Insight into

---

[9] This is the portfolio "rebalancing" effect first analyzed by Lance Girton and Dale Henderson.

[10] One might suspect from the above argument that the wealth elasticity of the demand for money is a crucial variable in determining initial period overshooting or undershooting. This can in fact be easily demonstrated by using instead of (1) a money demand function of the form:

$$m_t-p_t=-\delta r_t+\gamma(w_t-p_t)$$

Where γ is the wealth elasticity of the demand for money. With this demand function, it is straightforward to show that the exchange rate depreciates a maximum amount and in fact overshoots if $\gamma=0$; starting from $\gamma=0$, for larger and larger values of γ the amount of depreciation monotonically declines, becoming an undershoot at some $\gamma<1$.

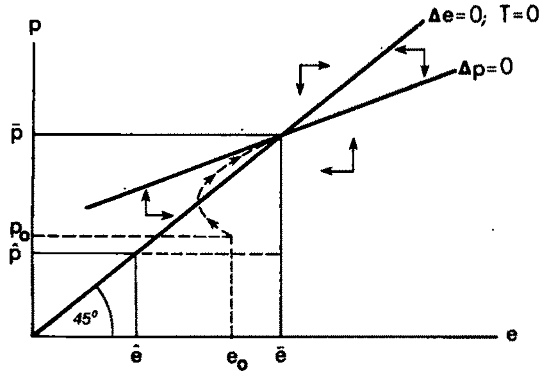


FIGURE 2. PHASE DIAGRAM WITH POSITIVELY SLOPED $\Delta e=0$ LINE

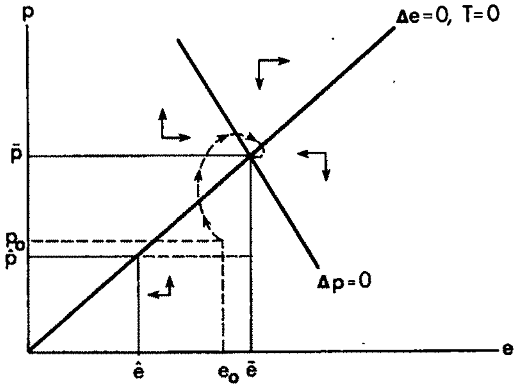


FIGURE 3. PHASE DIAGRAM WITH NEGATIVELY SLOPED $\Delta p=0$ LINE

their solution is obtained by studying their phase diagram in (e, p) space.

First consider equation (19). Setting Δe_t to zero, and solving for p_t yields

$$e_t=p_t \tag{21}$$

This 45° line through the origin may be called the exchange equilibrium line, and it coincides with the trade balance equilibrium or purchasing power parity line. To the left of this line, e_t is rising, and to the right e_t is falling.

The price equilibrium line is found by setting ΔP_t equal to zero in equation (20). Solving for P_{t-1} yields

$$p_{t-1}= \frac{-\lambda(1-\alpha)\theta(1+\theta+\beta)+\beta\theta(1-\alpha)}{-\lambda(1+\theta+\beta)+\beta\theta(1-\alpha)}e_{t-1}+N' \tag{22}$$

where N' is a constant. This line may be sloped positively or negatively. If positively sloped, however, its slope must be less than 45°, since $\lambda, \alpha,$ and θ are all fractions. The price equilibrium line also goes through the long-run equilibrium point $(\bar{e}, \bar{p})$.

The phase diagrams are graphed in Figures 2 and 3. Figure 2 presents the case where the slope of the price equilibrium line is positive, which occurs when β is larger than $\lambda(1+\theta+\beta)$. Figure 3 presents the case where the $\Delta p=0$ line is negatively sloped. Each case permits somewhat different paths.

One common aspect of the adjustment path is that it must cross the $(\Delta e_t=0, T=0)$ line at least once. Since the equilibrium value of f is a constant, we must begin and end with the same value of f. Surpluses in the trade balance must, at some point, be matched by deficits. Consider Figure 2. The exchange rate, following its instantaneous jump to e_0, begins appreciating as the trade balance runs a surplus and the home country accumulates foreign bonds. Eventually, the path must cross the $\Delta e_t=0$ line (in a vertical direction) and monotonically depreciate to its long-run equilibrium. Prices rise monotonically throughout the process.

Now consider Figure 3. In addition to a simple appreciation and consequent depreciation to long-run equilibrium, there is now the possibility of cycling around the equilibrium point.[11]

Which of these two permissible dynamic paths occur depends critically on the size of β, which measures trade balance responsiveness to relative price changes. There is always a large enough value of β to make the price equilibrium line positively sloped, given values of λ, θ, and α. The economic intuition is similar to that which explains why a large β constrains e_0 and p_0 to be close to the trade equilibrium line: if foreign exchange demands and/or supplies originating from trade flow considerations dominate demands arising from capital flow decisions, relative prices will be more in line with ones which equate the trade balance to zero. Note that, as $\beta \rightarrow \infty$, the adjustment path becomes the trade equilibrium line.

---

[11] Niehans (1977) also finds a possibility of cycling. Note that these dynamic patterns are also similar to those found in Niehans (1977) in that they have clockwise rotation. The characteristics of the equilibrium lines, though, are considerably different than the corresponding lines in Niehans' model. The differences appear to result from the assumption of period-by-period asset market equilibrium (for given values of wealth) in the present model as opposed to the Niehans' assumption of lagged asset adjustment.

V. Concluding Comments

Several points emerge from this analysis that cannot be understood or are not adequately emphasized in the context of the existing asset equilibrium models. First, the exchange rate is not determined solely in asset markets, nor solely in the foreign exchange or commodity markets, but simultaneously in all markets. While this in itself is perhaps well understood, the implications for exchange rate dynamics apparently are not. In particular, the characteristics of exchange rate adjustment paths when relative price effects are incorporated along with conditions of portfolio balance differ sharply from those predicted by the existing asset equilibrium approach. Furthermore, one of the striking implications of existing asset equilibrium models, that of initial overshooting, is seen to be not a general result, but one closely tied to rather special assumptions.

This analysis, of course, abstracts from many other presumably important considerations. One important omission has been noted in the text, that of a less than instantaneous adjustment of trade flows to relative prices. Incorporation of this would undoubtedly provide an alternative explanation for overshooting. Another deficiency is the restricted menu of assets in this model. By incorporating at least domestic government debt, the door would be opened for an analysis of alternative money supply mechanisms. A closely related point is that by incorporating variable income, the question of optimal policymaking involving managed floating could be analyzed. Finally, the rather crude modelling of expectations could be amended to incorporate perfect foresight.

REFERENCES

G. C. Archibald and R. G. Lipsey, "Monetary and Value Theory: A Critique of Lange and Patinkin," *Rev. Econ. Stud.*, Oct. 1958, *26*, 1–22.

G. A. Calvo and C. A. Rodriguez, "A Model of Exchange Rate Determination under Currency Substitution and Rational Expectations," *J. Polit. Econ.*, June 1977, *85*, 617–25.

R. Dornbusch, "Expectations and Exchange Rate Dynamics," *J. Polit. Econ.*, Dec. 1976, *84*, 1161–76.

L. Girton and D. Henderson, "Central Bank Operations in Foreign and Domestic Assets Under Fixed and Flexible Exchange Rates," in Peter B. Clark et al., eds., *The Effects of Exchange Rate Adjustments*, Washington 1974.

P. J. K. Kouri, "The Exchange Rate and the Balance of Payments in the Short Run and the Long Run: A Monetary Approach," *Scand. J. Econ.* No. 2, 1976, *78*, 280–304.

Jürg Niehans, "Exchange Rate Dynamics with Stock-Flow Interactions," *J. Polit. Econ.*, Dec. 1977, *85*, 1245–57.

______, *The Theory of Money*, Baltimore; London 1978.

D. Purvis, "Short-Run Dynamics in Models of Money and Growth," *Amer. Econ. Rev.*, Mar. 1973, 63, 12–23.

X-Inefficiency and Nonpecuniary Rewards in a Rent-Seeking Society: a Neglected Issue in the Property Rights Theory of the Firm

By W. MARK CRAIN AND ASGHAR ZARDKOOHI*

Several recent papers have stressed that opportunities to obtain monopoly profits will attract resources into efforts to obtain monopolies (see Gordon Tullock; Anne Krueger; Richard Posner). The major message of this literature is that traditional analysis and empirical estimates of the welfare losses from monopoly (see Arnold Harberger; David Schwartzman, 1960; David Kamerschen) may very substantially understate the total magnitude of these losses. However, the behavior underlying the tendency for monopoly rents to be transformed into genuine social costs—"rent-seeking activity"—has additional relevance for analysis of public vs. private enterprise. The presence of rent-seeking behavior results in predictable and important differences in the conduct and performance of public vs. private enterprise that are unexplained by the present property rights theory of the firm. While distinctions between public and private enterprise have been raised in the property rights literature, there remains no unambiguous prediction that privately owned firms will perform differently than public firms in a number of industries.[1] Indeed, there are at least two often-cited arguments which predict (particularly for utilities) that the conduct of firms in these two respective ownership categories will be similar. In this paper we argue that the bases of these apparent similarities between public and private enterprise are inapplicable in a rent-seeking environment, and that the introduction of this framework enhances the explanatory power of the property rights approach.

In Section I we briefly review the key arguments and offer in Section II an alternative characterization of public vs. private firm behavior in a world of rent seeking. In light of the implications of our analysis, Section III offers two types of empirical evidence on the contrasting incentives in public vs. private utilities. Specifically, we first compare the respective rewards structures for managers in publicly and privately owned water utilities in the United States. Second, we examine the impact of efficiency in private utility operation on rates of return subsequently allowed in formal rate hearings, under differing rent-seeking environments. Finally, Section IV offers some concluding remarks and discusses the relevance of our analysis for the welfare costs of monopoly.

I. The Managers of Utilities and the Utility of Managers

Modern treatments of enterprise operation and efficiency have given special emphasis to the internal cost-reward arrangements facing decision makers. While important conceptual differences pervade this literature, there are two basic lines of analysis which imply that managerial choices in publicly owned and privately owned utilities will not be much (if any) different. The first stresses the notion that all utilities—whether publicly or privately owned—generally operate in noncompetitive product markets. The second stresses that the institutional structure of utilities (for example, rate of return regulation) constrains owner access to, or personal use of, the pecuniary wealth of the enterprise. We briefly summarize these two arguments.

---

*Visiting assistant professor of economics, University of California-Los Angeles (on leave from Virginia Polytechnic Institute and State University), and Auburn University, respectively. We are grateful to James Buchanan, Harold Demsetz, Jack Hirshleifer, Albert Link, Robert Tollison, Gordon Tullock, participants at a workshop at the Hoover Institution, and an anonymous referee for comments on earlier drafts. Partial financial support for Crain's research time was provided by the Center for Study of Public Choice, Virginia Polytechnic Institute and State University.

[1] For two summaries of the property rights literature and empirical tests, see Eirik Furubotn and Svetozar Pejovich and our paper.

A. *X-Inefficiency*

The analytical tendency to treat all utility firms uniformly because of product market similarities traces to a seminal (although controversial) paper by Harvey Leibenstein (1966). The well-known X-inefficiency literature stresses the role of competitively organized product markets as an important source of disciplinary pressure on firm conduct and performance.[2] First, competition tends to eliminate high-cost producers, while the existence of monopoly power allows less efficient firms to remain in business. Second, competitive pressure squeezes profit to normal levels. This tends to bestir management and employees to put forth more effort and to motivate managers to combine inputs more efficiently. The Leibenstein analytical heritage therefore emphasizes the implication that utilities, irrespective of ownership form, are uniformly afflicted with X-inefficiency (higher operating costs) due to the general lack of pervasive pressure from competition in the product market.

B. *Constraints on Pecuniary Wealth*

Property right arrangements that attenuate owner access to the wealth of their enterprise (for example rate of return constraints) lower the cost of non-wealth-maximizing behavior on the part of managers. Gary Becker and Armen Alchian and Reuben Kessel offered the earliest formulations of this now familiar argument. In brief, this approach emphasizes that in institutional settings where "wealth cannot be taken out of an organization in salaries or in other forms of personal pecuniary property, the terms of trade between pecuniary wealth and nonpecuniary business-associated forms of satisfaction turn against the former" (Alchian and Kessel, p. 586). Predictable differences should thus arise between competitive vs. monopolistic (or profit constrained) firms. However, since owner access to firm wealth is generally attenuated for all utilities (regardless of ownership form), this argument also implies that non-wealth-maximizing (non-cost-minimizing) conduct should pervade regulated privately owned, as well as publicly owned, utilities.[3]

In sum, while the utility-maximization approach differs in important conceptual respects from the approach taken in the X-inefficiency literature, an important observable implication of the two frameworks does not.[4] Namely, all utilities, whether regulated private monopolies or publicly owned, have reduced incentives to minimize their cost of operation. Enterprise performance should predictably stray away from cost minimization because, (i) in the utility-maximization approach, cost-increasing production decisions do not come at the expense of owner-shareholder wealth, and (ii) in the X-inefficiency context, such decisions are not disciplined by pressures from alternative suppliers.

---

[2] Other papers which are pertinent to the development of the inefficiency argument include William Comanor and Leibenstein, Schwartzman (1973), Leibenstein (1973, 1975), and Walter Primeaux.

[3] We reiterate that distinctions have been developed in the property rights literature that do lead to predictable differences in public vs. private enterprise. For example, the high cost (if not the impossibility) of transferring ownership shares in publicly owned firms implies that owners have reduced incentives to monitor and to enforce cost-minimizing managerial behavior. See Alchian for the development and exposition of this point. He also stresses that the relatively greater diffusion of ownership shares, as in public enterprise, reduces the concentration of rewards and costs, and hence the dependency of an individual's wealth on his own efforts. These differences, however, are superceded in many industry settings (as argued in the text) in the presence of constraints on owner access to the wealth of private firms.

[4] A recent paper by George Stigler (1976) on the existence of X-inefficiency illustrates the conceptual gap. Stigler's message is essentially to stress the futility of the X-inefficiency concept for deriving welfare inferences. Managerial conduct which does not maximize the present value of an enterprise is not inefficient per se, but rather the efficient maximization of something else. (Robinson Crusoe, for instance, could never produce "inefficiently.") Stigler's objections, however, do not seem to be over the existence of non-wealth-maximizing behavior, but with the conceptual norm that such behavior is inefficient, since it is implausible that managers are inefficient in maximizing their utility.

II. The Relevance of Ownership Form in a Rent-Seeking Society

Rent-seeking activity adds another analytical dimension to the choice problem facing managerial decision makers. Where potential monopolies (regulated or otherwise) are available to privately owned firms, resources devoted to appropriating these property rights are simply a form of "investing" or bidding to acquire the monopoly rights. This bidding behavior would be affected, of course, by the expected payoff to the owners if such rights are obtained. Moreover, even where the current profits of a monopolist are explicitly regulated, for example, in private utilities, resources will be devoted to rate cases in seeking to influence the allowable rates of return. In effect, non-cost-minimizing behavior in regulated private utilities *does* come at the expense of owner wealth—or more precisely, at the expense of the *expected* wealth of owners. Non-cost minimization reduces the amount of resources that private firms have to expend in efforts to acquire, protect, or improve their present monopoly status. Therefore, the much-cited Alchian-Kessel postulate is not appropriate in a rent-seeking context; "[I]f regulated monopolists are able to earn more than the permissible pecuniary rate of return, then 'inefficiency' is a free good, because the alternative to inefficiency is the same pecuniary income and no 'inefficiency' " (p. 586). The opportunity cost of inefficiency is not zero, simply because the presence of potential monopoly profits creates an incentive to earn in excess of currently "permissible" pecuniary rates of return. Hence, nonpecuniary emoluments that enhance managerial utility will involve a tradeoff of owners' wealth. Owners of private utilities therefore have an incentive to monitor and to enforce cost-minimizing behavior on the part of managers even when constraints on current pecuniary profits are imposed.

With respect to the X-inefficiency argument, the rivalry to obtain monopoly production rights or efforts to raise regulated rates of return will exert competitive pressures—even in settings where there is only a single supplier during the actual production period. Indeed, in the limit, only those firms which can (and do) operate at minimum cost will achieve monopoly power in the rent-seeking context. Potential or present suppliers who suffer from X-inefficiency would be at a comparative disadvantage in the rivalry to acquire or maintain monopoly status. X-inefficiency could therefore not survive long in a rent-seeking society.

The potential to acquire, to protect, or to improve monopoly status thus gives importance to all pecuniary returns of privately owned enterprise, regulated or otherwise. On the other hand, the absence of rent-seeking opportunities in publicly owned firms means that the opportunity cost of inefficiency is comparatively lower. And while we have thus far developed this inference in the context of public vs. private utilities, it has broader applicability. For example, Alchian and Kessel argue that *all* monopolies are subject to constraints on pecuniary returns and, hence, that the distinction between explicitly regulated monopolies and unregulated private monopolies is a false one. If the profits in a private monopoly are large, it faces the threat of public policy sanctions or antitrust action. Alchian and Kessel argue that the potential for profit "regulation" of the antitrust variety creates roughly the same incentive structure in "unregulated" private monopolies as in monopolies (such as utilities) with explicit rate of return limitations.

Our analysis also suggests that privately owned monopolies, regulated or otherwise, will have roughly the same incentives, that is, rent-seeking activity is not unique to explicitly regulated private enterprise. For example, unregulated private monopolies can expend resources to influence public policy by supporting political candidates sympathetic to their company's interests, by swaying the determinants or outcomes of antitrust activity, or by influencing passage of legislation which protects their monopoly from potential entrants. Thus, in all privately owned monopolistic enterprises there is a potential return on resources expended to obtain or to protect monopoly profits, and this return represents the opportunity value of non-cost-minimizing operations.

In sum, the existence of rent-seeking provides an additional analytical basis for predicting differences in the behavior of public vs. private enterprise, even when both types of firms face similar product market conditions and/or restraints on "permissible" pecuniary profits. Rent-seeking activity provides a mechanism whereby monopoly rights will be allocated to those privately owned firms that can (and do) use them most "profitably," even where pecuniary constraints on owner wealth are ostensibly imposed during a current production period. This implies that the most productively efficient (least-cost) firms will acquire control of private monopoly rights through the allocative mechanism of rent seeking. While this mechanism tends to insure least-cost producers in the private sector—regulated or otherwise—the gain in production efficiency is offset by the welfare costs due to resources expended in the rent-seeking process, as stressed by Tullock, Krueger, and Posner. We return to this discussion of the net impact of rent seeking on the social costs of monopoly in the concluding section, but it is first appropriate to offer some evidence on our basic point—that rent-seeking activity is an important source of distinction between public and private enterprise.

III. Empirical Evidence

Differences in public and private enterprise have been observed empirically in several prior studies, although in many cases the theoretical underpinnings of these differences have been incorrectly specified.[5] We offer some additional evidence on the differential incentive structures in privately regulated vs. publicly owned firms that is implied by our analytical framework. Specifically, we first examine the hypothesis that owners in private firms (even though subject to state regulation) have greater incentives to induce cost minimization by managers than their counterparts in public firms. Second, we compare the relationship between private firm efficiency and subsequently allowed rates of return under two different legal settings which would affect the feasibility (i.e. costs) of rent-seeking behavior.

A.

One type of incentive mechanism which has observable merit is the extent to which managers are residual claimants to efficient (present-value-maximizing) performance of the enterprise (see Alchian and Harold Demsetz). In settings where owners have strong incentives to generate cost-minimizing managerial conduct, we would expect the reward structure to be such that managers can capture or internalize directly portions of the gains of achieving this objective. Thus, we would predict that where there are strong pressures from owners to minimize operating costs, managerial compensation would be tied to the performance of this goal.

A model to examine and compare the residual claimant mechanism in private vs. public firms is specified as follows:

$$(1)\quad ln\left(\frac{Y_i}{A_i}\right)=B_1 ln\left(\frac{1}{A_i}\right)+B_2 ln\left(\frac{C_i}{A_i}\right)+B_3 ln\left(\frac{R_i}{A_i}\right)+\mu_i$$

where Y_i = a proxy for total managerial compensation[6]

---

[5] For examples and discussion of the theoretical misspecification in previous empirical tests, see our paper, Crain and Robert Tollison, (1978); Louis DeAlessi, David Davies (1971; 1977), and Edwin Fugii and John Trapani.

[6] Data are not directly available on decision makers' incomes in our sample of water utilities. We chose administration cost in each firm as a proxy for managerial compensation. Administration cost for water utilities in the United States consists of approximately 70 percent managerial salary and about 30 percent office expenses such as secretarial salary and fringe benefits to managers such as travel expenses, etc. (The information was obtained from Mr. Craft of the American Water Works Association, Denver, Colorado.) Since managerial compensation could vary

A_i = total book value of the assets of firm i, net of depreciation[7]
C_i = average total operating costs of firm i, including maintenance, wages, and depreciation[8]
R_i = total sales revenues of firm i[9]
μ_i = a random disturbance term

This is the basic form of the model that we employ to examine the determinants of managerial compensation within a given type of enterprise. It follows the statistical procedure used by Wilbur Lewellen and Blaine Huntsman.[10] In order to make a statistical comparison of the incentive structures between public vs. private firms, we introduce a dummy variable, D_i, into equation (1) which is equal to *unity* if firm i is privately owned, and equal to *zero* if firm i is publicly owned. The D_i is used to check for both shifts and qualitative interaction effects with the other exogenous variables as specified in

$$(2)\quad ln\left(\frac{Y_i}{A_i}\right)=B_1 ln\left(\frac{1}{A_i}\right)+B_2 ln\left(\frac{C_i}{A_i}\right) + B_3 ln\left(\frac{R_i}{A_i}\right)+B_4 D_i ln\left(\frac{C_i}{A_i}\right) + B_5 D_i ln\left(\frac{R_i}{A_i}\right)+D_i+\mu_i$$

The results of estimating equation (2) using 1976 data on sixty public and fifty private water utilities are presented in equation (3). The T-ratios are in parentheses.

$$(3)\quad ln\left(\frac{Y_i}{A_i}\right)= \underset{(3.26)}{0.29}\; ln\left(\frac{1}{A_i}\right) - \underset{(-0.11)}{0.01}\; ln\left(\frac{C_i}{A_i}\right) + \underset{(9.85)}{0.78}\; ln\left(\frac{R_i}{A_i}\right) - \underset{(-2.33)}{0.13}\; D_i\left[ln\left(\frac{C_i}{A_i}\right)\right] - \underset{(-0.23)}{0.03}\; D_i\left[ln\left(\frac{R_i}{A_i}\right)\right] - \underset{(-2.70)}{1.53}\; D_i$$

No. observed = 110; $R^2 = 0.974$; $F_{(6,104)} =$ 662.23.

The findings presented in equation (3) indicate that in the privately owned firms managerial compensation is negatively and significantly related to operating costs. The estimated coefficient on C_i/A_i for the sample of private water utilities is significant at the .01 level. On the other hand, we observe no statistically significant relationship between operating costs and compensation for publically owned firms. Managerial renumeration in both types of water utilities does appear to be affected systematically by firm revenues. The estimated coefficient for R_i/A_i is positive and significant at the .01 level. There does not appear to be any difference in this relationship across the two

---

across firms for reasons such as differences in the cost of living in each city, we weight the compensation variable for each firm by the cost-of-living index for the area in which the firm is located. The weighting factor for each firm was derived using total expenditures (budget) of a "high-income" family of four on a "typical" market-basket of goods in the cost of living in each area where a firm is located. The data on the cost of living were obtained from U.S. Department of Labor, Table 30, pp. 276–77.

[7] Net book value was obtained from American Water Works Association.

[8] Average operating cost was determined by dividing total cost (including operating maintenance and depreciation costs) by the amount of water served to residential, commercial and industrial users. (See American Water Works Association.) Since a big portion of total cost is wages, and wages conceivably vary with the cost-of-living index, we weigh the average cost by the cost-of-living index as explained in fn. 6 above.

[9] Total sales revenues are the total receipts of each utility weighted by the cost-of-living index, as discussed in fn. 6. (See American Water Works Association.)

[10] Following Lewellen and Huntsman we divide each of the variables in equation (1) by assets *Ai* (the proxy for firm size) in order to eliminate problems of multicollinearity and heteroskedasticity. While Lewellen and Huntsman included measures of firm profitability in their empirical model of the determinants of managerial compensation, profitability is obviously not a useful parameter in our examination of publicly owned firms. Thus, our measure of operating costs *Ci* is substituted for profitability measures in order to check for the cost-minimizing determinant of managerial compensation.

types of firms, since the coefficient on $D_i[ln(R_i/A_i)]$ is not significant.[11]

This set of empirical results offers additional evidence on the different underlying conduct in public vs. private enterprise, even in an institutional setting where pecuniary profits in the private firms are constrained. The observable incentive structure facing managers in the private water utilities is one that rewards cost-minimizing behavior with higher managerial compensation packages. This is precisely the opposite type of incentive mechanism we would expect if owners did not want their enterprise to appear "too" profitable. Reward structures that induce cost-minimizing managerial behavior suggest that owners are indeed concerned about maximizing pecuniary returns, even though their access to current period profits are attenuated. Such returns in excess of current permissible rates may be useful in protecting or subsequently raising the permissible rates.

B.

The second type of empirical evidence that we offer concerns the relationship between operating efficiency in privately owned utilities and rates of return subsequently allowed by regulatory commissions in formal rate hearings. More specifically, we seek to examine further the pay off to regulated private firms (in terms of increases in allowed rates of return) from efficient performance. Our analysis suggests that private firms will have an incentive to operate efficiently in order to have more resources to devote to rent-seeking activity if these resources have a positive impact on future profits. In the specific context of rate-of-return regulatory settings, this implies that we should observe greater firm efficiency to be related positively to increases in allowed rates of return.[12]

We offer some additional evidence on the connection between firm efficiency and subsequent increases in allowed rates of return under two different policy settings. In some states the regulatory commissions permit regulated firms to make political contributions, to pay dues to trade associations for lobbying purposes, and/or to make expenditures for public relations in the rate-making process. In other states, however, the commissions have adopted policies that strictly disallow such activities directed at influencing the rate-making process by regulated firms (see U.S. Congress, p. 33).

We would thus expect the link between operating efficiency and the outcome of rate hearings to be more pronounced in those states where such expenditures are an accepted policy of regulatory commissions. This is not to say that rent-seeking expenditures will be eliminated in states where the commission does not legally allow regulated firms to devote resources for the purpose of influencing rate making. However, when these expenditures are illegal their prevalence should be deterred simply because of the added risk of sanctions or punishment.

The empirical model employed to examine these relationships is specified as follows:

$$(4)\quad ln(\Delta R_i) = b_0 + b_1 ln(E_i) + b_2 D_i ln(E_i) + b_3 ln(PC_i) + b_4 D_i + \mu_i$$

where ΔR_i = the difference between the old rate (i.e. for the test year) and the new allowed rate of return granted by the regulatory commission to firm i; E_i = proxy for efficiency in firm i during the test year, defined as the amount of net revenue, weighted by the permissible rates of return in effect during the test year, per dollar value of firm assets; PC_i = a proxy for projected cost in firm i, defined as the projected operating cost per dollar of "weighted" projected revenue. The weights are in terms of the new allowed rate of return. This weighting scheme is used because projected revenue in firm i is affected by the new allowed rate. The variable is used to control for the effect of expected differential increases in costs across firms; D_i = a binary variable which is equal to unity if firm i is allowed to make political expenditures in the rate making process and equal to zero otherwise; and μ_i = a random disturbance term.

---

[11] The relationship between a manager's compensation and the value of the wealth affected by his decisions is discussed in Alchian and William Allen (pp. 400, 795) and Crain and Tollison.

[12] In his study of regulation of gas and electric industries in New York, Paul Joskow finds tentative support for the hypothesis that "highly efficient firms are treated somewhat better by the commission once they enter the formal hearing process" (p. 641). Joskow, however, does not elaborate on the rationale underlying this observation. Moreover, his empirical measure of efficiency takes the form of an all-or-none dichotomy which is determined by the regulatory commission's subjective commendation of a particular utility's service.

We estimate the coefficients in equation (4) using data on thirty-four rate cases heard by State Regulatory Commissions for the two types of regulated water utilities in 1974.[13] The results of this estimation are presented as follows, t-ratios are in parentheses:

$$(5)\quad ln(\Delta R_i) = \underset{(-2.29)}{-19.81} - \underset{(-2.20)}{3.91}\, ln(E_i) + \underset{(2.34)}{4.18}\, D_i ln(E_i) + \underset{(2.53)}{0.94}\, ln(PC_i) + \underset{(2.39)}{20.5}(D_i)$$

No. observed = 34; $R^2 = 0.365$; $F_{(4,30)} = 4.17$.

The regression is significant at the .01 level, and the model "explains" roughly 36 percent of the variation in rates of return increases for the cross section of water utilities in 1974. The coefficient for the binary variable D_i is positive and significant at the .05 level, which supports the hypothesis that utilities experience greater increases in their allowed rates of return when rent-seeking expenditures are permitted by the State Regulatory Commissions. The coefficient on efficiency in states where utilities are allowed to make political expenditures in the rate-making process is positive and significant at the .05 level. This finding also offers evidence that greater operating efficiency increases the amount of funds which may be used to influence the outcome of formal rate hearings. For example, such funds could simply be used to hire better lawyers or more expert consultants. Finally, the coefficient on the variable controlling for expected cost increases PC_i is positive and significant at the .02 level.

As a final type of evidence on this point, we compare the mean of the differences between old rates of return (in the test year) and the rates of return granted by State Regulatory Commissions in 1974. The mean increase granted to utilities in states where the commissions's policy facilitates rent-seeking expenditures (i.e., where utilities are allowed to make political expenditures in the rate-making process) is 4.19 percent, and the mean increase in states that have adopted policies that inhibit rent-seeking expenditures is 2.53 percent. We can reject the hypothesis that these two means are equal at the .05 level in favor of the alternative hypothesis that rate increases are higher in those states with the former type of policy setting. This result offers additional support for our argument that inefficiency is not a "free good" even in rate-of-return regulated private utilities, because non-cost-minimizing behavior will predictably reduce the expected increases in rates of return. Moreover, in policy settings that more openly permit rent-seeking expenditures by firms, the relationship between operating efficiency and subsequently allowed profits is of a greater magnitude.

IV. Concluding Remarks: The Social Costs of Monopoly

The major purpose of this paper has been to stress the relevance of rent-seeking activity in expanding the predictive power of the property rights theory of the firm. The potential for private firms to influence their expected future profitability with rent-seeking expenditures means that observable differences should arise between public and private enterprise behavior even under apparently similar product market conditions. We note, however, that there is additional relevance for evaluating the social costs of monopoly and regulation. First, we can predict that what managers may lose in terms of less work effort (i.e., X-inefficiency) or nonpecuniary returns (i.e., utility maximization), political entrepreneurs may gain through rent-seeking activity. That is, one consequence of rent seeking is that owners have incentives to redistribute wealth toward those individuals with political power to grant monopoly status. At one level this effect only represents a redistribution of wealth from managers in the private sector to those in the political sector. At a secondary level, however, it trades one sort of inefficiency for another, which tends to reinforce the case put forward by Tullock, Krueger, and Posner that rent-seeking activity represents a welfare cost from monopoly which exceeds the deadweight losses due to monopoly quantity restriction. Specifically, the incentive that we have described would tend to keep down production or "standard" operating costs in monopoly, which would consequently raise the level of rent-seeking expenditures above what they would be in the presence of X-inefficiency or nonpecuniary rewards.

---

[13] The sources of the data on utility operations and allowed rates of return is National Association of Regulatory Utility Commissioners (1975; 1976). The data for the commission policies regarding expenditures vis-à-vis the ratemaking process is U.S. Congress.

REFERENCES

Armen A. Alchian, "Some Economics of Property Rights," *Il Politico*, Dec. 1965, 30, 816–29.

______**and William R. Allen,** *University Economics*, 2d ed., Belmont 1967.

______**and H. Demsetz,** "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*, 777–95.

______**and R. E. Kessel,** "Competition, Monopoly, and the Pursuit of Money," in H. Gregg Lewis, ed., *Aspects of Labor Economics*, Princeton 1962, 157–75.

Gary S. Becker, *The Economics of Discrimination*, Chicago 1957.

W. S. Comanor and H. Leibenstein, "Allocative Efficiency, X-Efficiency and the Measurement of Welfare Losses," *Economica*, Aug. 1969, *36*, 304–09.

W. M. Crain and R. D. Tollison, "State Budget Sizes and the Marginal Productivity of Governors," *Public Choice*, Fall 1976, *27*, 91–96.

______**and**______, "On the Sufficient Conditions for Wealth Maximizing Conduct," *Kyklos*, No. 3, 1978, *31*, 500–03.

______**and A. Zardkoohi,** "A Test of the Property Rights Theory of the Firm," *J. Law. Econ.*, Oct. 1978, *21*, 395–408.

D. G. Davies, "The Efficiency of Public Versus Private Firms: The Case of Australia's Two Airlines," *J. Law. Econ.*, Apr. 1971, *14*, 149–65.

______, "Property Rights and Economic Efficiency: The Australian Airlines Revisited," *J. Law. Econ.*, Apr. 1977, *20*, 223–26.

L. De Alessi, "Managerial Tenure under Private and Government Ownership in the Electric Power Industry," *J. Polit. Econ.*, May/June 1974, *82*, 645–53.

E. T. Fujii and J. M. Trapani, "On Estimating the Relationship between Discrimination and Market Structure," *Southern Econ. J.*, Jan. 1978, *44*, 550–67.

E. G. Furubotn and S. Pejovich, "Property Rights and Economic Theory: A Survey of Recent Literature," *J. Econ. Lit.*, Dec. 1972, *80*, 1137–62.

A. C. Harberger, "Monopoly and Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1954, *44*, 77–87.

P. L. Joskow, "The Determination of the Allowed Rate of Return in a Formal Regulatory Hearing," *Bell J. Econ.*, Autumn 1972, *3*, 632–44.

D. R. Kamerschen, "An Estimation of the 'Welfare Losses' from Monopoly in the American Economy," *Western Econ. J.*, Summer 1966, *4*, 221–36.

A. O. Krueger, "The Political Economy of the Rent-Seeking Society," *Amer. Econ. Rev.*, June 1974, *64*, 291–303.

H. Leibenstein, "Allocative Efficiency vs. X-Efficiency," *Amer. Econ. Rev.*, June 1966, *56*, 392–415.

______, "Competition and X-Efficiency: Reply," *J. Polit. Econ.*, May/June 1973, *81*, 765–77.

______, "Aspects of the X-Efficiency Theory of the Firm," *Bell J. Econ.*, Autumn 1975, *6*, 580–606.

W. G. Lewellen and B. Huntsman, "Managerial

Pay and Corporate Performance," *Amer. Econ. Rev.*, Sept. 1970, *60*, 710–20.

R. A. Posner, "The Social Costs of Monopoly and Regulation," *J. Polit. Econ.*, Aug. 1975, *83*, 807–27.

W. J. Primeaux, "An Assessment of X-Efficiency Gained through Competition," *Rev. Econ. Statist.*, May 1977, *59*, 105–08.

D. Schwartzman, "The Burden of Monopoly," *J. Polit. Econ.*, Dec. 1960, *68*, 627–30.

______, "Competition and Efficiency: Comment," *J. Polit. Econ.*, May/June 1973, *81*, 756–64.

G. J. Stigler, "The Xistence of X-Efficiency," *Amer. Econ. Rev.*, Mar. 1976, *66*, 213–16.

______, "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Econ. J.*, June 1967, *5*, 224–32.

G. Tullock, "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Econ. J.*, June 1967, *5*, 224–32.

American Water Works Association, *Operating Data for Water Utilities*, AWWA No. 20112, Denver 1976.

U.S. Congress, Senate Committee on Government Operations, *State Utility Commissions*, Washington 1967.

U.S. Department of Labor, "Annual Costs of a Higher Budget for a 4-Person Family," Washington 1976.

Comparing Utility Functions in Efficiency Terms: Comment

By Dana N. Stevens and James E. Foster*

In a recent note in this *Review*, Burton Weisbrod argues that in those cases where "consumption possibilities are not independent of preferences," it may be possible to compare different "types" of utility functions on the basis of economic efficiency. Weisbrod suggests a "base reversal" test where

> ...one type of utility function, and the expected consumption bundle it generates, may be said to be preferred to another, and the expected consumption bundle it generates, if and only if (a) the two expected consumption bundles are different and (b) the same consumption bundle is preferred no matter which utility function is used to evaluate the two bundles. [p. 993]

This test may indeed yield binary rankings for different types of utility functions.[1] Unfortunately, these binary rankings may be cyclic if more than two types of utility functions are examined. While it is possible to compare utility functions by employing the Weisbrod test, such comparisons may have a limited usefulness for policy decisions.

Let X be the set of all *possible* choice alternatives under consideration. For a particular utility type U, only some of the possible alternatives are actually *feasible* or *attainable*; the subset A of X comprises these elements. If the choice alternatives are consumption bundles, then consumption possibilities will depend upon the underlying type of utility function.[2] To complete the model, let C be the relation which associates with each pair (U, A) the set of choice alternatives that can be expected to result from the existing economic and political choice processes. The general Weisbrod test may then be stated as follows: U is preferred to U' if and only if, for all alternatives c in $C(U, A)$ and all c' in $C(U', A')$, $c\ P_i c'$ and $c\ P_i' c'$ for every individual i in society.

Consider Table 1 in which three types of utility functions are compared. The set of possible alternatives is $X = \{x, y, z\}$. In our example, a Type I utility function is clearly preferred to a Type II utility function, since in either case all individuals prefer the outcome y from adopting Type I to the outcome z from adopting Type II. Similarly, Type II is preferred to Type III. But Type III is preferred to Type I, generating a cyclic ranking. None of the three utility types can be considered unambiguously best.[3]

The possibility of a cyclic ranking does not by itself limit the usefulness of applying the Weisbrod test. After all, the Voting Paradox (from which our example is derived) is well known; yet no one suggests that democratic processes be revoked be-

---

*Assistant professor of economics, New College, University of South Florida; and graduate student, Cornell University, respectively.

[1] In general, a type of utility function may be considered any possible configuration, or n-tuple $(R_1, \ldots, R_N)$, of individual preference orderings, irrespective of whether these orderings actually can be represented by real-valued utility functions.

[2] Weisbrod considers shifts in the set of feasible alternatives to be related to the ability of some types of utility functions to "internalize" externalities without the overt expenditure of productive resources. However, the phenomenon may be more common than Weisbrod suspects. For example, dropping the assumption of fixed factor supplies, the amount of labor supplied to the market may reflect different preferences for work and leisure. Different types of utility functions will generate different factor supply elasticities, thus affecting consumption possibilities.

[3] The distinction between possible and feasible sets of alternatives is not needed if the choice process can be Pareto inefficient. With inefficient choice processes, the outcomes of our example can be associated directly with choice over X. In either case, the Weisbrod test yields a cyclic ranking. We are grateful to the referee for suggesting the possibility of inefficient choice.

TABLE 1

Utility Type	Feasible Set	Choice Outcome
U_{I}: For all i, xP_iyP_iz	$A_{\mathrm{I}}=\{y, z\}$	$C(U_{\mathrm{I}}, A_{\mathrm{I}})=\{y\}$
U_{II}: For all i, $y\ P_iz\ P_ix$	$A_{\mathrm{II}}=\{z, x\}$	$C(U_{\mathrm{II}}, A_{\mathrm{II}})=\{z\}$
U_{III}: For all i, $z\ P_ix\ P_iy$	$A_{\mathrm{III}}=\{x, y]$	$C(U_{\mathrm{III}}, A_{\mathrm{III}})=\{x\}$

cause of a possible cyclical majority. The difference is that the Voting Paradox *is* well known. It has been examined carefully, both in theory and in practice. We can recognize those preferences that are likely to generate the paradox. And we have some idea how to resolve the paradox if it arises in actual voting situations.

In contrast, we know very little about the "efficiency" properties of different utility types. We lack empirical data on the nature of the attainable sets associated with the various types of utility functions; we need information on how attainable sets will change when preferences are manipulated. We lack empirical data on the actual (nonmarket) processes of collective choice; we must discover how the political system itself will be amended as a result of changing preferences. Lacking such empirical substance for the Weisbrod test, it seems premature "to devote resources to shaping utility functions."

REFERENCE

B. A. Weisbrod, "Comparing Utility Functions in Efficiency Terms," *Amer. Econ. Rev.*, Dec. 1977, *67*, 991–95.

Comparing Utility Functions in Efficiency Terms: Reply

By BURTON A. WEISBROD*

If utility functions can be compared in efficiency terms, as I argued is sometimes the case (see my earlier paper), it follows that there is a new research area for economists to explore. To be sure, no analytic perspective, no matter how useful it may be, is without problems. For example, as Dana Stevens and James Foster have correctly pointed out, the possibility (though not the certainty) of cyclic ranking of utility functions exists if more than two types are compared. Their comment is, I believe, intended not as criticism of the utility function comparability perspective but as a call for more research; for example, on the efficiency properties of various types of utility functions, and on the responsiveness of attainable commodity sets to changes in preferences.

The point is that there do exist researchable questions that are suggested as soon as we recognize the possibility of subjecting utility functions to analysis within a Paretian framework. Conventionally such questions have not been asked, partly because they were regarded as equity matters on which economists had little to say normatively, and partly because the factors shaping utility functions have been regarded as lying outside the domain of economics.

It certainly is premature to say that we now know enough to conclude that on efficiency grounds alone, resources should— or should not —be devoted to shaping utility functions. Yet to entertain even the possibility that one type of utility function can be "preferred" to another, and that changing such a function *may* be efficient, is to reach out in a bold new direction. Important conceptual issues as well as vital public policy questions are at stake.

REFERENCES

D. N. Stevens and J. E. Foster, "Comparing Utility Functions in Efficiency Terms: Comment," *Amer. Econ. Rev.*, Sept. 1980, *70*, 793–94.

B. A. Weisbrod, "Comparing Utility Functions in Efficiency Terms," *Amer. Econ. Rev.*, Dec. 1977, *67*, 991–95.

---

*University of Wisconsin-Madison.

Bank Credit and the Labor-Managed Firm: Comment

By FRANK H. STEPHEN*

The economics of the labor-managed firm is a field of study which has grown in prominence over the last twenty years. Important papers on this topic have appeared in the *Review*, for example, by Benjamin Ward, Evsey Domar, Jaroslav Vanek (1969), and most recently by Erik Furubotn (1976). The earlier papers were concerned with modelling the short-run decision process of such a firm and were clearly exploratory in nature. Furubotn's paper, however, is concerned with long-run decision making and is possibly the high point of a series of papers presenting a critique of the property-rights structure implicit in one form[1] of labor management. The contribution of Furubotn and his collaborator Svetozar Pejovich is sufficiently distinctive that, given their institutional affiliations, they will be referred to below as the Texas School.

In common with most of the writings of the Texas School, Furubotn presents a model of a labor-managed firm which has no recourse to credit to finance its purchases of productive assets. This is somewhat surprising since the school of writers who take their inspiration from the work of Vanek lay great emphasis on external financing. Furthermore the institutional framework implicit in Furubotn and explicit in most of the other papers from the Texas School is that of Yugoslavia which exhibits a very high degree of credit financing of investment. To some extent this is explicable in that three papers which explicitly deal with the impact of credit finance (see Furubotn and Pejovich, pp. 280–81; Pejovich, Furubotn, 1974) conclude that the availability of credit merely exacerbates the problems which arise in the self-financed case. This view may well be taking on the mantle of conventional wisdom in that it has been repeated by Branko Horvat, J. Michael Montias, and to some extent Laura Tyson, and one of the papers is itself from an important collection of essays in the field of comparative systems edited by Morris Bornstein. The objective of the present paper is to demonstrate that this conclusion is subject to considerable qualification.

The analysis leading to the Texas School's conclusion on credit is as follows. They argue that the existence of bank credit reduces voluntary savings within enterprises and that consequently there will be a retardation of voluntary saving in an economy composed solely of such firms. It will be presently argued that the former is not necessarily true, and that even when it is, the reduction in self-finance is less than the analysis of the Texas School suggests, and finally that any retardation of voluntary savings arises from an unwillingness of the banking authorities to allow the rate of interest to be set by a competitive market process.

In analyzing the consequences of the availability of credit, the Texas School employs a rule of "credit first," that is, a collective will only self-finance investment if prof-

---

*Senior lecturer in economics, University of Strathclyde, Glasgow. I would like to thank James Love for his comments on this note and the managing editor and referee of the *Review* for their comments on an earlier draft written while I was a visiting fellow at the Program on Participation and Labor-Managed Systems, Cornell University. The financial assistance of the Carnegie Trust for the Universities of Scotland and the British Academy are acknowledged.

[1] Although Furubotn's paper is an analysis of what might be called the genus "labor management," it deals exclusively with the Yugoslav species of that genus. Other species which exhibit different behavior are Mondragon (see Anne Johnson and William Whyte), the Scott-Bader Commonwealth (see Ernest Bader), and the theoretical model of Vanek (1970). It may also be useful to distinguish between "socialist" labor management (for example, the models of Furubotn and Vanek) where property is socially owned and "private" labor management (for example, the model of K. V. Berman and M. D. Berman) where individual members of the collective retain a private claim on the returns from investment.

itable investments remain after all available credit has been exhausted. Furubotn states that this "rigid sequence is made provided certain assumptions are accepted" (1974, p. 269). These assumptions are: (a) the marginal product of capital falls monotonically, (b) the supply of bank credit is infinitely elastic at the going rate, (c) the rate of interest is low enough relative to the return on capital to permit some positive borrowing, (d) the rate of time substitution of consumption increases steadily as current consumption is diminished, and (e) the collective is concerned exclusively with maximizing a utility function whose only arguments are individual consumption in each period of the planning horizon. It can be demonstrated using Figure 1, which is adapted from Furubotn (1974), that the conclusion regarding this "rigid sequence" is wrong.

In this figure the horizontal axis measures levels of savings and investment undertaken by the collective with a specified planning horizon. It is assumed that all members of the collective have identical preferences and any distribution of income is egalitarian. The vertical axis measures the rate of interest paid on savings and the return from investment in physical assets.

The *II* schedule represents the marginal efficiency of investment schedule for the enterprise and *SS* is the savings schedule for members of the collective when saving takes the form of deposits with the banking system. When internally financed investment takes place under the property-rights structure assumed by the Texas School, the worker-managers lose all claim to the capital invested which they are obliged to maintain via compulsory depreciation payments. Thus the current income forgone can only be regained from the net return on such an investment. Consequently, internally financed investment projects must yield a net return considerably higher than that offered on bank deposits for worker-managers to be indifferent between the two.[2] At a given rate of interest on bank deposits (for example, i_o), the required return on internally financed investment within the firm will be higher (for example, r^*). When the bank deposit rate is i_o, no corporate saving to finance investment will take place if the net return on it is below r^*. For higher rates on bank deposits, correspondingly higher rates of return on internally financed investment will be required for worker-managers to be indifferent between the two forms of investment. Thus with a bank deposit rate of i_o, the supply of "corporate" savings to finance investment in physical assets is given by r^*BS'. The savings schedule of the collective shifts because the property-rights structure drives a wedge between the productivity of an asset to the firm and the return to the members of the firm. This is what Roger McCain has called the "Furubotn-Pejovich Effect."

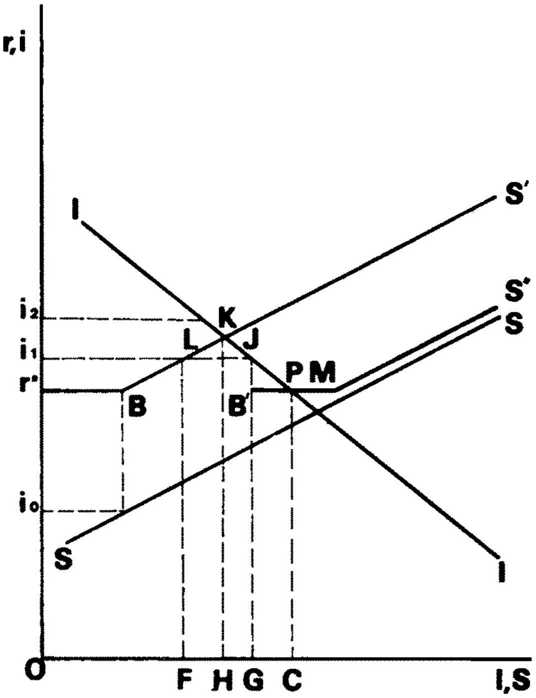


FIGURE 1

If i_1 is the rate at which the collective may borrow funds, Furubotn's "rigid sequence" would lead to investment of *0C* of which *0G* is borrowed and *GC* internally financed. The level of borrowing is given by the level of investment on which the marginal return is i_1 (i.e., $0G$). The level of corporate saving and therefore internal finance is given by the investment, in excess of that financed through borrowing, whose marginal return lies between r^* and i_1 (i.e., GC). The determination of the level of internally financed investment is equivalent to a rightward shift of the corporate savings schedule (r^*BS') by the amount of borrowing undertaken. The corporate savings schedule when $0G$ is borrowed is shown in the figure by $B'MS''$.

---

[2] A fuller justification of this is given, for example, in Furubotn (1974, 1976), Furubotn and Pejovich, Pejovich, and Vanek (1971).

Intuitively Furubotn's sequence seems illogical since it implies utilizing the resources with the higher opportunity cost (i_1) in preference to those with the lower (r^*). Analytically, in terms of Marshallian surpluses, Furubotn's rule implies sacrificing the area $r^* BLi_1$ in order to obtain the area $B'PJ$. The latter cannot, under the assumptions made by Furubotn, be greater than the former. Thus the availability of credit will not substantially reduce the volume of self-finance—in the absence of credit the level of investment would have been $0H$. It would seem that Furubotn is led into this error by a concern with maximizing the level of investment.

Thus far all that has been demonstrated is that the application of the seemingly more rational "cheapest source first" rule means that the introduction of credit reduces self-financing by a lesser extent than that suggested by the Texas School. However it is possible that circumstances will arise where no credit is taken up by an individual enterprise even though the rate of interest is below the rate of return on some projects. If credit were available at the rate of i_2 in the figure, the cheapest source first rule would dictate that all investment should be self-financed. This would be the case whenever the rate of interest was greater than KH. It must therefore be concluded that the Texas School is wrong to assert that the availability of credit reduces self-finance. It would also be wrong to conclude that the availability of credit necessarily increases an enterprise's level of investment above its wholly self-financed level.

The configuration of schedules in Figure 1 does of course indicate that the level of investment given labor management will be less than under private property rights. However, this is a matter of construction: if the savings schedules were such that SS intersected i_1J to the left of J, when i_1 is the price of credit, then under both systems the level of investment would be $0G$. In such a case the labor-managed firm would have a greater demand for credit.

The application of what has been argued above to be the economically rational rule for evaluating credit by the labor-managed firm has results which are not really surprising. The rule itself is derived from conventional financial theory and was utilized, for example, by Jack Hirschleifer. The analysis still, however, indicates that the availability of credit *may* reduce self-finance and the Texas School clearly feels that in an economy composed solely of labor-managed firms, the consequences are of more significance than under private property rights. They suggest that it results in "a tendency toward a general retardation of voluntary saving in the economy" (Furubotn and Pejovich, p. 282). This will only be so if the interest paid by the banks to their depositors is insufficient to bring forth the level of savings necessary to meet the demand from potential borrowers, or if there is some other form of interference in the market for loanable funds. The Texas School's conclusion is a function of their assumptions about the workings of the banking system and not the property-rights structure. This may be valid for the particular case of Yugoslavia, but not for a generalized treatment such as that of Furubotn's paper.

REFERENCES

E. Bader, "From Profit Sharing to Common Ownership," *J. Curr. Soc. Issues*, No. 1, 1971, *10*, 4–15.

K. V. Berman and M. D. Berman, "The Long-Run Analysis of the Labor-Managed Firm: Comment," *Amer. Econ. Rev.*, Sept. 1978, *68*, 701–05.

Morris Bornstein, *Plan and Market, Economic Reform in Eastern Europe*, New Haven; London 1973.

E. D. Domar, "The Soviet Collective Farm as a Producer Cooperative," *Amer. Econ. Rev.*, Sept. 1966, *56*, 734–57.

E. G. Furubotn, "Bank Credit and the Labor-Managed Firm: The Yugoslav Case," in his and Svetozar Pejovich, eds., *The Economics of Property Rights*, Cambridge Mass. 1974, ch. 18.

______, "The Long-Run Analysis of the Labor-Managed Firm: An Alternative Interpretation," *Amer. Econ. Rev.*, Mar. 1976, *66*, 104–23.

______ **and S. Pejovich,** "Property Rights, Economic Decentralization, and the Evolution of the Yugoslav Firm: 1965–72," *J. Law Econ.*, Oct. 1973, *16*, 275–302.

J. Hirschleifer, "On Optimal Investment Decision," *J. Polit. Econ.*, Aug. 1958, *66*, 329–52.

Branko Horvat, *The Yugoslav Economic System*, White Plains 1976.

A. G. Johnson and W. F. Whyte "The Mondragon System of Worker Production Cooperative," *Ind. Labor Relat. Rev.*, Oct. 1977, *31*, 18–30.

R. A. McCain "On the Optimum Financial Environment for Worker Cooperatives," *Z. Nationalökon.*, No. 3–4, 1977, *37*, 355–84.

J. Michael Montias, *The Structure of Economic Systems*, New Haven; London 1976.

S. Pejovich, "The Banking System and Investment Behaviour of the Yugoslav Firm," in Morris Bornstein, ed., *Plan and Market, Economic Reform in Eastern Europe*, New Haven; London 1973.

L. D'A. Tyson, "A Permanent Income Hypothesis for the Yugoslav Firm," *Economica*, Aug. 1977, *44*, 393–408.

Jaroslav Vanek, "Decentralization under Workers' Management: A Theoretical Appraisal," *Amer. Econ. Rev.*, Dec. 1969, *59*, 1006–14.

______, *The General Theory of Labor-Managed Market Economies*, Ithaca 1970.

______, "The Basic Theory of Financing of Participatory Firms," in his *The Labor-Managed Economy*, Ithaca; London 1977, ch. 9.

B. Ward, "The Firm in Illyria: Market Syndicalism," *Amer. Econ. Rev.*, Sept. 1958, *48*, 566–89.

Bank Credit and the Labor-Managed Firm: Reply

By EIRIK G. FURUBOTN*

Frank Stephen's comment on the theoretical views of the "Texas School" proceeds at two levels; he raises some general questions concerning the applicability of the property-rights approach to comparative economic analysis, and he attempts to show that certain specific errors appear when the new theory is employed to explain the demand of the socialist labor-managed firm for bank credit. Both levels of criticism require response. Because of the limitations of space, however, it will be necessary to confine discussion to the relatively well-defined technical issues associated with the problem of bank credit.

The argument that Yugoslav banking policy, rather than the property-rights structure, explains the behavior of the labor-managed firm is developed by Stephen in terms of four analytical cases based on his Figure 1. Collectively, these cases are supposed to show that the Texas School is wrong in its conclusions about the collective's savings habits and use of bank credit. What must be determined, though, is whether the criticisms can themselves stand up to close analysis.

I

Stephen's first two cases deal with the firm's reaction to particular credit charges (i_1 and i_2, respectively), given a savings bank rate of i_o and certain other background assumptions.[1] The important thing to note is that these cases are *special*; they relate to circumstances in which the bank rate charged for credit is set at an abnormally high level. In each of the illustrations, the cost of credit (i_1, i_2) is greater than the critical rate of return (r^*) required for voluntary investment in nonowned assets. But in what must be considered the typical situation under labor management, the bank rate is almost certain to be *less* than r^*. Unless the collective's planning horizon (T) is very long (and the savings rate i_o quite low), the adjustment for the property-rights structure must lead to a large value for the critical rate r^* and, thus, r^* is likely to be greater than any bank rate of interest (i) reflecting the general opportunity cost of capital.[2] The quantitative relations here are suggested by Table 1 in which the values for r^* are derived for various possible horizons (T) on the assumption that the rate i_o paid on savings deposits (owned assets) is 5 percent.

If the firm's effective planning horizon is ten years, the real rate of return (r^*) that must hold in order for the collective to be willing to undertake self-financed investment is about 13 percent. The assumptions of Stephen's first two cases, of course, demand that the cost of credit be even higher. But interest rates of such magnitude would not be imposed unless the government were deliberately seeking to prevent the use of bank credit.[3] In the more reasonable case illustrated in my Figure 1, the cost of credit i_3 is less than r^*. Then, the firm will depend exclusively on bank credit for capital accumulation and total investment will be *0v*.

Obviously, the magnitudes of the parameters i_o and T are crucial to the financial solution reached. The position of my paper is that while the rate paid on savings de-

---

*Professor of economics, Texas A&M University.

[1] These assumptions do not accord completely with the Yugoslav institutional structure, see my forthcoming paper. The latter model considers the firm's demand for bank credit under conditions in which the firm's planning horizon is less than the period of the bank loan.

[2] Stephen's first two cases have their best chance for occurrence when i_o is low relative to the opportunity cost of capital and i_1 is high. In his fourth case however, Stephen seems to argue that when labor management functions normally the savings bank rate (i_o) will equal the cost of credit (i_1). But if this condition holds, his first and second cases will never appear.

[3] Presumably, by one means or another, the government can always *force* firms to undertake some self-financed investment.

TABLE 1[a]

T	r^*	r^{**}
1	105.	108.
5	23.1	25.0
8	15.5	17.4
10	12.9	14.9
20	8.0	10.2

[a]Shown in percent.

posits may or may not be set at a relatively low level, the firm's effective horizon is likely to be short. The reasons for the latter conclusion can be sketched as follows. In considering self-financed investment, workers must recognize that the firm's policies are decided within the organization through some sort of democratic political process (see my 1976 paper). But, over time, the membership of the collective may change, and change in such a way as to produce new voting patterns, revised payoff policies, etc. Thus, to estimate his probable payments stream, a worker must estimate how long the existing political coalition controlling the firm will endure, and how long current policies will continue to be enforced. Given the uncertainties of firm dynamics, the perceived horizon is likely to be, at best, only moderately long. Moreover, even if no changes occur in the membership of the collective and the initial horizon (T_o) is lengthy, all workers will age as time elapses and retirement is approached. Consequently, at any cross section of time t, the effective horizon (T_t) is given by the formula: $T_t = T_o - t, t = 1, 2, \ldots, T_o$. Clearly T_t does not have to be lengthy (see my 1979 paper).

Despite the argument just presented, it is still conceivable that the expected horizon can be substantial (for example, $T=20$). At the same time, Stephen is correct in asserting that the collective's net income (inclusive of rents to worker-investors) is at a maximum when the rule "cheapest source first" is followed in financing the firm's investment. But do these facts imply that workers will never attempt to substitute bank credit for their own savings when, for example, $i_2 > r^*$? The answer is no and the reason lies in the uncertainty surrounding reward for self-financed investment. Workers who invest in the firm have no special legal claim on the firm's revenues.[4] While the enterprise has a *contractual obligation* to repay a bank loan, it need never reward worker-investors with higher wages. Moreover, the longer the horizon, the greater the potential risk. Thus, if risk-averse workers wish to maximize *expected utility*, they may find it expedient to forego self-finance and sacrifice some rent. In effect, "insurance" against a deficient income stream can be purchased. By using bank credit and maximizing the level of the firm's investment, workers are able to protect themselves against loss, preserve their industrial mobility, and achieve modest diversification in their "investment portfolios" (see James Meade, pp. 426–27, and Jacques Drèze).

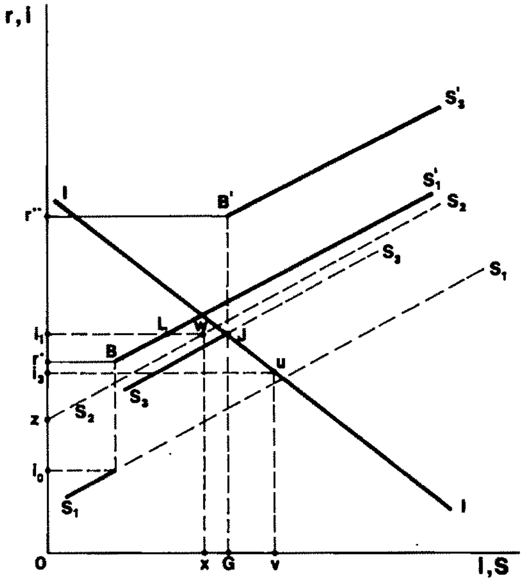


FIGURE 1

II

The third case presented by Stephen is designed to show that the property-rights structure underlying socialist labor management exercises little independent effect on behavior and, in particular, does not prevent the labor-managed firm from undertaking as much investment as an orthodox capitalist firm operating subject to private property rights. In geometric terms, the case is based on an assumed shift of the savings schedule from S_1S_1 to S_2S_2 (in my Figure 1). Properly interpreted, the shift implies that there will be *no saving whatsoever* by the workers of the labor-managed system. That is, if the savings bank rate remains at i_o (as tacitly suggested), investment in owned assets will be nil (i.e., $z > i_o$). At the same time, if the firm's planning horizon does not lengthen significantly, the required rate for investment in nonowned assets must lie above i_1; and hence, the firm will use bank credit *exclusively* to finance its investment volume of $0G$. Stephen's discussion seems to imply that self-financed investment will account for $0x$ dollars of the total investment but this interpretation is wrong; the "uncorrected" savings schedule S_2S_2 merely indicates the willingness of workers to save and invest in *owned* assets at different possible rates of interest.

---

[4] The calculation of r^* is based on the assumption that *all* of the increased profit attributable to capital expansion will (with certainty) be paid out each period in the form of wage increases. If only a fraction of the productivity improvement is devoted to raises, the original calculation is, of course, compromised.

The consequences of labor management just noted seem sufficiently distinctive to warrant attention. Certainly, it is misleading to say that the two systems of business organization produce the same results. In principle, the bank rate (i) under labor management can be adjusted so as to yield an investment solution that mimics the capitalist solution. But, for the case considered, an important difference still remains: No savings are made by socialist workers and the labor-managed firm's capital accumulation is financed entirely by bank credit.

III

The theme of the final case presented by Stephen is that retardation of voluntary saving in a labor-managed economy will occur only if the interest rate paid by savings banks to their depositors is held at an artificially low level, or if other interference with the "capital market" exists. On this interpretation, the property-rights structure plays no role in the problem of retardation. The argument here is, in fact, an extension of Stephen's third case and can be explained as follows. Assume that, in Figure 1, S_3S_3 intersects II at point J. Assume further that the rate i_1 now has a dual meaning; it represents both the reward paid on savings deposits (owned assets) and the going rate charged for bank credit. Since the return on owned assets has been raised from i_o to the higher level of i_1, the minimum required rate of return on nonowned assets, say r^{**}, must be correspondingly higher. My Table 1 suggests how this conversion is made when the savings bank rate goes from 5 to 8 percent. In general, once corrections have been made and the new savings schedule $r^{**}B'S_3'$ has been established, the collective's attitude toward self-financed investment can be understood. Conditions are such that the cost of credit i_1 is less than the critical rate r^{**}. Thus, given freedom, the firm will use only bank credit to finance its investment ($0G$). The situation is interesting because, if i_o is allowed to float up to i_1, the cost of credit will lie below the critical rate no matter what the length of the collective's planning horizon.[5]

Effectively, then, self-financed investment in the labor-managed firm is ruled out if it is required that the savings bank rate be brought into equality with the commercial bank rate. It also follows, of course, that if this arrangement of interest rates is regarded as normal and appropriate for labor management, Stephen's first and second cases lose all practical significance. Bank financing of investment is not merely important, it is universal.

The preceding analysis may seem puzzling because, given my assumptions, the amount workers save ($0G$) is precisely equal to the amount the firm invests. But this equality does not mean that the firm's investment is truly "self-financed." The terms of the respective financial transactions have to be kept in mind. Workers save $0G$

---

[5] As long as the collective's horizon is *finite*, as it must be under the institutional structure assumed, the condition $r^{**} > i_1$ holds.

and invest in *owned* assets; in effect, they have a legal agreement with a savings bank by which they receive a definite rate of interest (i_1) on their deposits and have the freedom to withdraw principal and interest at any time. On the other side, the firm, as such, undertakes a contractual obligation to a commercial bank to repay the investment loan in accordance with some established schedule. What must be recognized here is that, under the conditions of Stephen's fourth case, incentives are such as to dissuade workers from investing in nonowned assets. Indeed, workers are completely cut off from "equity" investment.

Ceteris paribus, it is this limitation on investment choice that diminishes the general willingness of individuals to save in a labor-managed economy. From the standpoint of any decision maker, legal investment alternatives promising higher rewards than the fixed rate of interest paid on savings deposits (i_1) are not available even if the individual wishes to take substantial risks.[6] A varied array of financial instruments is lacking and no comprehensive system of differentiated capital markets exists. Thus, relative to the situation under capitalism, it is more difficult to mobilize voluntary savings. It is also true that as long as bank financing of investment is, de facto, the only method used, workers will find it harder to see the connection between personal saving and the progress of their own firm. Saving will not seem to be a necessary precondition to capital formation and jobs. Indeed, if bank loans are forthcoming,[7] workers can seem to have the best of both worlds; they can share in any residual their firm makes and, at the same time, avoid sacrificing consumption or risking their own funds in nonowned investment. Under these general circumstances, then, the position of the savings schedule in a labor-managed system is likely to be to the left of its capitalist counterpart. The point is important because what we are concerned with in the analysis is not just movement up and down a *given* savings schedule (as Stephen's discussion suggests) but the relative position of the schedule.

The implications of this line of discussion are clear. Not only is retardation of voluntary saving likely to occur in a labor-managed system, but the explanation of the phenomenon is found in the property-rights structure extant. Moreover, it is this same property structure that accounts for freely chosen financial plans that exclude completely, or severely limit, self-financed investment by the collective. There is no doubt that troublesome questions still remain concerning the best method for financing the labor-managed firm[8] (see Drèze, and Schlicht and von Weizsäcker); nevertheless, it would seem that property-rights analysis must play a role in the development of any improved theory.

---

[6] Each worker, of course, takes certain risks by virtue of his membership in a labor-managed firm. He participates in the firm's residual in accordance with definite rules laid down by the collective (for example equal sharing). (See Ekkehart Schlicht and C. C. von Weizsäcker and Michael Jensen and William Meckling.) The individual also has the option of investing in human capital.

[7] In theory, the banking system is supposed to coordinate the savings and investment decisions of the system and channel the savings made by workers (in the savings banks) to the firms via loans. If voluntary savings are modest, however, it seems likely that the government will react by putting excessive credit into the system, thus generating inflation.

[8] Since all true labormanaged systems share certain common characteristics, similar forms of property-rights analysis are useful in considering the problems of labor-management in the socialist setting (see my 1976 paper) and the capitalist setting (see my 1978 paper).

REFERENCES

J. H. Drèze, "Some Theory of Labor Management and Participation," *Econometrica*, Nov. 1976, *44*, 1125–39.

E. G. Furubotn, "The Long-Run Analysis of the Labor-Managed Firm: An Alternative Interpretation," *Amer. Econ. Rev.*, Mar. 1976, *66*, 104–23.

______, "The Long-Run Analysis of the Labor-Managed Firm: Reply," *Amer. Econ. Rev.*, Sept. 1978, *68*, 706–09.

______, "Decision Making Under Labor Management: The Commitment Mechanism Reconsidered," *Z. ges. Staatswiss.*, June 1979, *135*, 216–27.

Income, Labor Supply, and Urban Residence

By JOHN S. HEKMAN*

The framework for urban spatial models as developed in the last twenty years contains two main features: access and bid rent. In a recent article in this *Review*, William Wheaton (1977b) has done serious damage to this framework by concluding that the long-run spatial equilibrium theory derived from bid rent "contributes little to the explanation of American location-income patterns" (p. 631). He finds that the theory developed by William Alonso, Richard Muth, and others produces a very weak tendency for middle- and upper-income families to live at a greater distance than poor families from the central business district (*CBD*). The purpose of this note is to argue that the Alonso-Muth theory is an important tool in understanding urban structure if account is taken of one previously ignored factor—that the labor force participation of married women varies dramatically and inversely with the income of their husbands. Middle- and lower-income households have more of their members in the labor force, and this pulls these groups closer to the *CBD*.

Muth and Wheaton have noted that there is only a weak relationship empirically between median family income and distance from the *CBD*. Muth (p. 263) found that income did not increase with distance on the South Side of Chicago when the age of the housing stock was included in the regression. This gives more support to the filtering model of urban structure than to Muth's analysis.[1] Wheaton (1977b) finds that bid-rent schedules decrease very little in slope as income increases, thus producing a very weak tendency for higher-income households to live farther out.

The models underlying the analyses of Wheaton and Muth rely on the simplifying assumption of a one-worker household. A caveat is usually stated regarding the effect of multiple-worker households (see Muth, p. 41; Wheaton, 1977a, p. 206). Others have looked at the two-worker household but have not made the connection between working wives and husbands' incomes (see Walter Oi; Michelle White; Oded Hochman and Haim Ofek), although this connection was noted some time ago in the labor economics literature (see Glen Cain). The fact that wives' labor supply is a major factor in the urban labor force is illustrated by Table 1. The participation rates of married women, husband present, decline with husband's income in almost every category, except the numerically unimportant case of husbands who earned under $3,000. In most categories the rates at the $3,000–6,000 level are double or more the rates for the over $15,000 level. However, participation rates also differ quite a bit by age, both because of life cycle effects such as schooling and childbearing, and because of the pronounced time trend of wives' participation, shown in Table 2. Younger labor force members have made different decisions regarding careers and investment in human capital, which argues for treating them separately from older workers.

The participation of married women in the labor force has been increasing steadily for about a century. This coincides with Wheaton's observation on the timing of the movement of upper-income groups to the suburbs, which, he says, "has been evolving gradually ever since the streetcar suburbs of the late nineteenth century" (1977b, p. 620). Thus the changing characteristics of the labor force have accompanied the changing location of households by income.

---

*Boston College. I received a great deal of helpful criticism on this paper from Christine Hekman, Marvin Kraus, John Meyer, and Joseph Quinn. Financial support for this research was provided by the Harvard-M.I.T. Joint Center for Urban Studies.

[1] I am indebted to John Meyer for valuable help in pointing out some of the implications of Muth's empirical results.

Table 1—Full-Time Labor Force Participation Rates of Married Women, Husband Present, by Age of Woman and Husband's Income, 1970

Age of Wife	Husband's Income Under $3,000	$3,000–5,999	$6,000–9,999	$10,000–14,999	Over $15,000
			I. Central Cities of *SMSA*s		
16–21	.296	.300	.309	.250	.134
22–24	.431	.371	.336	.268	.250
25–34	.340	.328	.288	.195	.134
35–44	.350	.368	.343	.251	.152
45–64	.304	.346	.325	.267	.177
			II. Other Urban		
16–21	.272	.277	.290	.241	.181
22–24	.415	.369	.306	.229	.201
25–34	.305	.317	.258	.152	.094
35–44	.331	.390	.327	.225	.125
45–64	.296	.358	.329	.264	.168

Source: See U.S. Bureau of the Census, *Employment Status and the Work Experience*, Table 16.

Table 2—Female Labor Force Participation Rates by Marital Status, 1890–1970

	> 16-Years Old					> 15-Years Old			
	1890 (1)	1900 (2)	1910[a] (3)	1920 (4)	1930 (5)	1940 (6)	1950 (7)	1960 (8)	1970 (9)
Total	18.9	20.6	25.4	23.7	24.8	25.8	29.0	34.5	42.6
Married	4.6	5.6	10.7	9.0	11.7	13.8	21.6	30.7	40.8
Single	40.5	43.5	51.1	46.4	50.5	45.5	46.3	42.9	53.0

Sources: Various U.S. Censuses. This table is adapted from Claudia Goldin.

[a] The 1910 labor force figures are too high relative to those for other years because the Census reported certain women employed in agriculture as in the labor force rather than at home.

The inclusion of labor supply effects in the bid-rent analysis alters Wheaton's conclusion in that there is an additional income effect to be calculated. In the Alonso-Muth approach, Wheaton says, "If land is more income elastic than total marginal travel costs (*MCT*), then greater income reduces the slope of the bid price gradient, suggesting that the wealthy will live further from the center" (1977b, p. 622). Wheaton's condition for the wealthy to live farther out is basically the same as those of Muth (p. 30) and Edwin Mills (p. 87) except for the treatment of travel costs. Wheaton, following Alonso, includes the disutility of travel in the utility function, while Muth treats the cost of travel as lost leisure time. It is worth noting here that Muth's analysis is in error in one respect. The utility function he uses contains q, the quantity of housing services, and x, which is "all commodities except housing and transportation but including leisure" (thus the variable y called income by Muth also includes the value of leisure time). The price of x is normalized at $p_x = 1$ and the price of housing services is p. Muth goes on to derive first-order conditions for spatial equilibrium; he then examines the effect of a change in income on this equilibrium. However, a change in wage income will violate the normalizing restriction $p_x = 1$, since x includes leisure time, valued at the wage rate. There will be substitution of goods, including housing, for leisure, and this is not included in Muth's derivation of the income-distance relationship. The effects of the value of time in this context have been worked out by Hochman and

Ofek; in addition, Mills manages to avoid making Muth's mistake and still obtains the same equilibrium condition.

Muth's analysis can be amended rather easily to take account of family labor supply. Let the utility function be written as

$$U = U(x, q, l_f, l_m) \tag{1}$$

where the ls are the leisure time of the wife and husband, respectively, and x is all commodities except housing and transportation. To this must be added time constraints for the husband and wife, and an income constraint, where y is earnings alone:

$$\begin{aligned} L_m + l_m + bk &= H \\ L_f + l_f + g(L_f)k &= H \\ y = w_m L_m + w_f L_f& \\ = x + p(k)q + T(k, L_f)& \end{aligned} \tag{2}$$

Here H is total hours available for work, commuting and leisure; k is distance from the *CBD* in miles; L is hours spent working; and T ( ) is the money cost of commuting. The husband is assumed to make a fixed number of work trips and to have a fixed per mile time cost of travel, b. For wives there is far more variation in days, hours, and weeks worked; $g(L_f)$ is the function relating the time spent commuting to labor supply, L_f.[2]

Substituting the time constraints into the income constraint and forming a Lagrangian expression, the result is

$$\begin{aligned} L = U(x, q, l_f, l_m) + \lambda\big[& w_m(H - l_m - bk) \\ & + w_f(H - l_f - g(L_f)k) - x \\ & - p(k)q - T(k, L_f)\big] \end{aligned}$$

The equilibrium bid-rent schedule derived from L is

$$\left.\frac{\partial L}{\partial k}\right|_{x, q, l_f, l_m, \lambda} = -\lambda\left[g(L_f)w_f + kw_f\frac{dg}{dL_f}\frac{\partial L_f}{\partial k} + bw_m + \frac{\partial T}{\partial k} + q\frac{dp}{dk}\right] = 0$$

or

$$\frac{dp}{dk} = -\frac{g(L_f)w_f + kw_f\dfrac{dg}{dL_f}\dfrac{\partial L_f}{\partial k} + bw_m + \dfrac{\partial T}{\partial k}}{q} \tag{3}$$

where is λ is the Lagrange multiplier associated with the combined income and time constraints. Muth's corresponding expression is

$$\frac{dp}{dk} = -\frac{\partial T/\partial k}{q} \tag{4}$$

where $T = T(k, y)$ for Muth, thus including the cost of time. Expression (3) differs mainly in making the slope of the bid-rent schedule dependent on the wife's labor supply and on wage rates rather than earnings. This latter consideration avoids the error of treating y as the measure of family welfare. This error results because an increase in w_m is associated with a decrease in L_f; to the extent that the wife's labor supply decreases, the use of y results in an underestimate of the change in the real welfare of the family. The proper view of the wife's reduction in market labor is that she has substituted more time at home as the value of marginal product of her husband's time at work rises.

Equation (3) is the negative of the ratio of marginal per mile transportation cost to the quantity of housing consumed. Following Muth, q is the demand for housing, but in this instance it is $q = q(w_f, w_m, p)$, whereas for Muth, $q = q(y, p)$. The use of wage rates w_f and w_m is superior to earnings y because wage rates are directly proportional to the resource constraint concept of family welfare developed by Gary Becker; an earnings variable used as the income measure in a demand function will not capture differences among families in the market vs. nonmarket allocation of time.

---

[2] Alternatively, L_f can be thought of as the proportion of women in full-time employment at a particular level of aggregation such as the census tract or, as with the data used here, the community.

If we assume, with Muth and Wheaton, that $\partial T/\partial k$ is a constant, namely, that dollar per mile transportation costs are invariant with distance, then the numerator of (3) is a function of the variables k, l_f, w_f, and w_m, so that (3) can be written in general form as

$$\frac{dp}{dk}=-\frac{t(k,L_f,w_f,w_m)}{q(w_f,w_m,p)} \tag{5}$$

where $t(\ )$ is per mile transportation cost for the household. In this context, dp/dk can be considered the slope of an indifference curve in (p,k) space. High-income families (where income is measured by wages) will live farther from the center if an increase in income causes the slope of $p(k)$ to become flatter. To see this, the percentage change form of (5) can be used:

$$\left(\frac{d\hat{p}}{dk}\right)=-(\hat{t}-\hat{q})=-\big[E(t,w_f)\hat{w}_f + E(t,w_m)\hat{w}_m+E(t,L_f)\hat{L}_f+E(t,k)\hat{k} - E(q,w_m)\hat{w}_m-E(q,w_f)\hat{w}_f-E(q,p)\hat{p}\big] \tag{6}$$

where a "hat" denotes a percentage change and $E(\ )$ is an elasticity. Following Muth and Wheaton, we are looking at the change in the slope of $p(k)$ at a point, so $dk=0$; then $E(t,k)\hat{k}=0$ and $E(q,p)\hat{p}$ is also zero because $dp=(dp/dk)dk$. Then (6) can be rearranged to read

$$\left(\frac{d\hat{p}}{dk}\right)=\big[E(q,w_m)\hat{w}_m+E(q,w_f)\hat{w}_f - E(t,w_m)\hat{w}_m-E(t,w_f)\hat{w}_f\big] - E(t,L_f)\hat{L}_f \tag{7}$$

The bracketed term in (7) is the income-distance condition discussed by Muth (p. 30), Mills (p. 87), and Wheaton (1977b). This term alone says that high-income households will live at a greater distance from the urban center if the income elasticity of housing demand (represented by wage rates in (7)) exceeds the elasticity of transportation cost with income.

Going beyond the bracketed term, equation (7) says that changes in wives' labor supply will act through transportation cost to affect the slope of bid rent. A simple model of the determinants of L_f will be introduced below; the point here is that Wheaton's empirical result which shows the bracketed term to be close to zero does not mean that bid rent is not having an effect on the household's location decision. The slope of bid rent can become flatter with increases in income even when the bracketed term is zero if $E(t,L_f)\hat{L}_f$ is negative.

The labor supply term in (7) is made up of two components: the elasticity of per mile commuting cost with respect to labor supply and the change in labor supplied. The first of these is unambiguously positive, but the sign of the second will depend on the determinants of L_f. Studies such as those of Cain and Cain and Michael Dooley have used w_m, w_f, asset income, and various age and fertility variables to estimate wives' labor supply. Table 3 presents estimates of such a supply function for a single metropolitan area, using observations for eighty-two communities of over 10,000 population in the Chicago Standard Metropolitan Statistical Area (*SMSA*), including the central city. The "in the labor force" participation rates are multiplied by the proportion of working women who worked fifty weeks or more, in order to approximate more closely a full-time rate. Since married women, husband present, are not broken down by age, the supply function is also reported for all women in the 25–34 and 35–44 age groups. Equation 1 in Table 3 shows that male income explains most of the variance of labor force participation rates. In the other estimates, the signs of most coefficients are as expected, except for that of w_f for women 25–34. There is a spurious correlation between *LFPR* and w_f because the variable used is earnings, not the wage rate.

TABLE 3—ESTIMATES OF THE DETERMINANTS OF LABOR FORCE PARTICIPATION RATES (*LFPR*) FOR WOMEN IN THE CHICAGO *SMSA*

Dependent Variable	Estimated Coefficients				
	Constant	*ln* w_m	*ln* w_f	*ln* child	*ln k*
1. L: *log LFPR* for married women	13.27	−1.13			
husband present	(17.4)	(−13.8)			
$R^2=.70$; $F=190.6$					
2. L_1: same as 1	3.84	−0.85	0.80	−.005	0.03
$R^2=.78$; $F=69.8$	(1.6)	(−9.4)	(4.4)	(−0.04)	(0.5)
3. L_2: *log LFPR* for	18.63	−0.97	−0.02	−1.05	0.009
women 25–34	(6.1)	(−8.6)	(−0.1)	(−6.7)	(0.1)
$R^2=.78$; $F=67.2$					
4. L_3: *log LFPR* for	6.28	−.93	0.69	−0.05	−.03
women 35–44	(2.5)	(−9.9)	(3.0)	(−0.4)	(−0.5)
$R^2=.80$; $F=75.5$					

Source: U.S. Bureau of the Census, *General Social and Economic Characteristics*.

Note: Variables: w_m = median earnings of males; w_f = median earnings of females; child = percent of women who have children under 17; k = airline miles from *CBD*. The *t*-statistics are in parentheses beneath estimated coefficients.

This is only partly adjusted for by the full-time participation rate variable.

The distance variable k comes in with little or no effect on participation in Table 3. This suggests that distance (transportation cost) is not an important enough factor to influence the decision to participate in the labor force, although whether one participates is quite important in choosing a location. The number of women in the labor force decreases markedly with distance from the center, as the simple regressions of labor supply on distance show:

$$L_1 = \underset{(11.4)}{4.35} - \underset{(-4.2)}{0.316}\ ln\,k \quad R^2=.18;\ F=18.0$$

$$L_2 = \underset{(12.8)}{5.40} - \underset{(-6.4)}{0.525}\ ln\,k \quad R^2=.34;\ F=40.9$$

$$L_3 = \underset{(12.4)}{4.88} - \underset{(-5.0)}{0.383}\ ln\,k \quad R^2=.24;\ F=24.8$$

The range of female participation rates ("in the labor force, aged 25–39"), taken from individual Chicago *SMSA* census tracts, is from 0.93 at one-half mile from the corner of State and Madison to 0.18 at twenty-eight-miles distant.

The labor supply estimates from Table 3 can be used with equation (7) to shed more light on the income-distance relationship. L_f will change when the wage rates of husband and wife change as follows:

$$(8) \quad \hat{L}_f = E(L_f, w_m)\hat{w}_m + E(L_f, w_f)\hat{w}_f$$

With this addition, equation (7) becomes

$$(9) \quad \left(\frac{d\hat{p}}{dk}\right) = \big[E(q, w_m)\hat{w}_m + E(q, w_f)\hat{w}_f$$
$$- E(t, w_m)\hat{w}_m - E(t, w_f)\hat{w}_f\big]$$
$$- E(t, L_f)\big[E(L_f, w_m)\hat{w}_m + E(L_f, w_f)\hat{w}_f\big]$$

The sign and magnitude of the labor supply term depends on the relative increases in w_m and w_f. There is no "representative" division of changes between w_m and w_f, but husbands' wages vary a great deal more than those of wives, so the likely range of $\hat{w}_m$ and $\hat{w}_f$ is from $\hat{w}_f = 0$ to $\hat{w}_f = w_m$. Using the estimates of $E(L_f, w_f)$ and $E(L_f, w_m)$ from Table 3, equation 2, the range of the labor supply term in (10) would be 0.05 $E(t, L_f)$ to 0.85 $E(t, L_f)$. The magnitude of $E(t, L_f)$, the

TABLE 4—INCOME-DISTANCE REGRESSIONS FOR THE CHICAGO *SMSA*, 1970
(Dependent Variable = *Log* of Miles from the *CBD*)

	Log of Median Family Income	L_1	L_2	L_3	Age of Housing[a]	R^2	F
1.	.44					.07	5.7
	(2.3)						
2.	−.39	−.82				.21	10.3
	(−1.4)	(−3.7)					
3.	−.26		−.75			.36	21.8
	(−1.3)		(−6.0)				
4.	−.72			−1.04		.30	16.8
	(−2.6)			(−5.1)			
5.	−.19	−.67			−.08	.43	19.9
	(−.78)	(−3.6)			(−5.6)		

[a] Age of housing = percent of housing stock constructed in 1939 or earlier.

elasticity of the household's per mile transportation cost with the wife's labor supply, depends on what type of work schedule is assumed. More hours worked per day with the same number of work trips would not increase transportation cost at all. But in the more frequent case of women working more days or weeks (or, using aggregated data, when more women enter the labor force), the elasticity would be about unity times the wife's share of transportation cost. Since the wife's share is bounded by 0 and 1, $E(t, L_f)$ has a range from 0 to 1, and the entire labor supply term will range from 0 to 0.85. Since this term is added to the Alonso-Muth-Wheaton elasticity condition, it means that the income elasticity of housing demand can be less than the elasticity of transportation cost with income by a considerable margin while the bid-rent curve still becomes flatter at higher incomes.

The possibility that, holding labor supply constant, the wealthy have a tendency to live closer rather than farther from the center is supported by the evidence that in the nineteenth century the income ordering in American cities was the reverse of that existing today (see, for example, Homer Hoyt). The positive correlation of median family income and distance from the *CBD* today is rather weak, casual observation notwithstanding. Muth found that this correlation disappeared when he included the age of housing in his regression. However, his sample was from the South Side of Chicago in 1950, when that area was in the process of becoming the country's largest black community by the mechanism of neighborhood turnover. This hardly seems the ideal test of the filtering model of urban structure. The sample of communities in the Chicago *SMSA* used for the regressions in Table 3 also shows a very weak income-distance relationship. This is reported in equation 1, Table 4. The addition of female labor force participation variables causes the sign of the income-distance relationship to be reversed, although the coefficient is only significant in one case out of three. When the age of the housing stock is included in equation 5, the labor supply variable retains its significance, while the income relationship becomes weaker. Thus the income-distance relationship is definitely misspecified when *LFPR* is not included. The income effect on the bid-rent schedule cannot be separated from the labor supply effect.

This paper has attempted to demonstrate that the bid-rent approach to urban structure has not been discredited by Wheaton's study. Household labor supply is a crucial variable in any model of urban structure and it varies in a way which is much more systematic than previous writers have assumed. The urban labor force today differs from that of the past in that more work is done outside the home by married women. Lower-income households have more labor force participants (including children) than higher-income ones. The income-distance relationship cannot be tested properly unless the income-labor supply interaction is included, and the results reported here indicate that the rich might actually live closer in were it not for differences in labor supply among income groups.

REFERENCES

G. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

Glen Cain, *Married Women in the Labor Force*, Chicago 1966.

______ **and M. Dooley,** "Estimation of a Model of Labor Supply, Fertility and Wages of Married Women," *J. Polit. Econ.*, Aug. 1976, *84*, S179–S200.

G. Carliner, "Income Elasticity and Housing Demand," *Rev. Econ. Statist*, July 1973, *55*, 528–32.

C. Goldin, "Female Labor Force Participation: The Origin of Black and White Differences, 1870 and 1880," *J. Econ. Hist.*, Mar. 1977, *37*, 87–108.

O. Hochman and H. Ofek, "The Value of Time in Consumption and Residential Location in an Urban Setting," *Amer. Econ. Rev.*, Dec. 1977, *67*, 996–1003.

Homer Hoyt, *One Hundred Years of Land Values in Chicago*, Chicago 1933.

Edwin Mills, *Urban Economics*, Glenview 1972.

Richard Muth, *Cities and Housing*, Chicago 1969.

W. Oi, "Residential Location and Labor Supply," *J. Polit. Econ.* Aug. 1976, *84*, S221–S238.

A. M. Polinsky, "The Demand for Housing: A Study in Specification and Grouping," *Econometrica*, Mar. 1977, *45*, 447–61.

W. Wheaton, (1977a) "A Bid-Rent Approach to Housing Demand," *J. Urban Econ.*, Apr. 1977, *4*, 200–17.

______, (1977b) "Income and Urban Residence: An Analysis of Consumer Demand for Location," *Amer. Econ. Rev.*, Sept. 1977, *67*, 620–31.

M. White, "A Model of Residential Location Choice and Commuting by Men and Women Workers," *J. Reg. Sci.*, Apr. 1977, *17*, 41–52.

U.S. Bureau of the Census, *U.S. Census of Population: 1970, Employment Status and the Work Experience*, Subject Report PC(2)-6A, Washington 1973.

______, *U.S. Census of Population: 1970, General Social and Economic Characteristics*, Vol. I, Part 15, Washington 1973.

In Defense of Some "Paradoxes" of Trade Theory

By HORST HERBERG AND MURRAY C. KEMP*

In recent years there has been a systematic exploration of the implications of relaxing the strict assumptions of the textbook version of the Heckscher-Ohlin theory of international trade. Some of the best-known propositions have turned out to be highly vulnerable to changes in assumptions. Thus if the assumption of constant returns to scale is relaxed, one must be prepared to sacrifice the Stolper-Samuelson and Rybczynski theorems (see Ronald Jones, 1968, and Kemp), and one must accept the possibility that commodity prices and outputs are negatively associated (see our 1969 paper). And if factor-market distortions are admitted, one must be prepared for the possibility that the ranking of industries by value factor intensities may be opposite to the ranking by physical factor intensities, implying in turn that outputs and prices may be negatively associated and that the Stolper-Samuelson and Rybczynski theorems must be rephrased and then cease to be dual to each other (see our paper with Stephen Magee; Jones, 1971). Against the background of traditional results, the new findings have been viewed as "perverse" and "paradoxical."

Now a reaction has set in. Wolfgang Mayer and J. Peter Neary have argued that the paradoxes of static neo-Heckscher-Ohlin models are incompatible with the stability of certain dynamic models so that, if the assumption of stability is granted, the paradoxes will almost never be observed.

Our purpose in this comment is to suggest that the dynamic analysis carried out by Mayer and Neary is inconclusive. We wish to make one basic point, equally applicable to both papers, and then to illustrate the point with reference to Neary's paper. (Mayer's paper would have served equally well.)

Stolper, Samuelson, and Rybczynski, as well as Jones and the present authors in the cited papers, were concerned with the comparative static properties of one part of a larger general equilibrium system. For purposes of examining those properties, it was quite proper to abstract from the rest of the system by artificially freezing commodity prices and factor endowments. However, in searching for stability conditions to impose on comparative static calculations within the smaller system, one must look at a dynamic extension not of the smaller but of the larger static system. There seems to be no good reason to insist that any particular part of a larger dynamic system should be stable in isolation. Suppose that we have built up a larger static system by adding commodity demand and factor supply functions, and that we have dynamized the system and calculated stability conditions. Those conditions may be useful in comparative static computations within the smaller model. But one would not expect to find that stability of the smaller dynamic model is implied by stability of the larger dynamic system. It is true that the postulate of stability under all speeds of adjustment, together with other mild assumptions, does imply that all subsystems are stable. However it does not seem that either Neary or Mayer had that much stronger assumption in mind.[1]

We now turn to a detailed study of Neary's paper. He considers a small price-taking country which produces two products with two primary factors. At each moment of time, the allocation of factors to industries is given "by history." In each industry at each moment producers pay for each factor the value of its marginal product in that

---

*University of Kiel and University of New South Wales, respectively.

[1] Other points could have been made. For example, every static model can be imbedded in any number of different dynamic extensions, and each dynamic extension has its own stability conditions. Conclusions drawn from just one or two such extensions therefore must be highly tentative. However, this point is well known.

industry. Depending on the absence or presence of factor-market distortions, a factor may or may not receive the same reward in both industries. Even in the absence of permanent distortions, factors may receive rentals which vary from industry to industry ("transient" factor-market distortions). Over time each factor tends to move from the industry in which it is relatively poorly rewarded to the industry in which it is relatively well rewarded.

Neary shows that such an economy is globally stable or unstable as the ranking of the two industries at rest is the same or different in terms of relative value and physical factor intensities, that is, in Neary's notation, as $|\theta|\cdot|\lambda|>0$ or $|\theta|\cdot|\lambda|<0$.[2] He concludes that stationary equilibria in which relative physical and value factor intensities differ will almost never be observed and that the comparative statical paradoxes such as perverse price-output relationships based on such a difference can be consigned to the category of "theoretical curiosa."[3]

Neary has performed a useful service in reminding us that the relevance of comparative static calculations can be established only by stability analysis of an appropriate dynamical system. It is not clear, however, that his own analysis is appropriate. Neary's dynamic model is partial equilibrium in the sense that both commodity prices and factor endowments are taken to be unexplained constants. It therefore can be imbedded in a more general model in which prices and endowments are endogenous. And the more general model may be stable when Neary's model is unstable.

It remains to show that there do exist interesting extensions of Neary's model and that these more general models may be stable even when Neary's model is unstable. The first such generalization is provided by Neary himself. After developing his main proposition (in Sections I and II) he goes on (in Section III) to close his model by introducing homothetic preferences, the same for each individual, and price-determining commodity markets which clear at each moment. He shows that the extended model is locally stable if and only if

$$(1)\quad \sigma:=\sigma_D|\theta|\cdot|\lambda|+\sigma_X Q_X+\sigma_Y Q_Y>0$$

where σ_D is the elasticity of substitution in demand, σ_i is the elasticity of factor substitution in the ith industry, and

$$Q_i:=\lambda_{Li}\theta_{Ki}+\lambda_{Ki}\theta_{Li}\quad(i=X,\,Y)$$

Obviously condition (1) is satisfied if

$$(2a)\qquad |\theta|\cdot|\lambda|\geqslant 0$$

Alternatively, as can easily be shown,[4] condition (1) holds if

$$(2b)\qquad \sigma_D\leqslant\sigma_X+\sigma_Y$$

In particular, (2b) implies (1) even if $|\theta|\cdot|\lambda|<0$, that is, even if Neary's sufficient condition for fixed-price instability is satisfied.

One can go a step beyond Neary's Section III and extend his flexible-price model by treating one of the two commodities as an investment good and thus allowing for capital accumulation. Such a model is in turn a special case of one developed in our 1972 paper. Specifically, it is obtained from our model by making all savings propensities equal to each other and constant. From

---

[2] $|\lambda|$ and $|\theta|$ are the determinants of the matrices of factor-input coefficients λ_{ij} and factor-cost shares θ_{ij}, respectively ($i=L,K;\ j=X,Y$).

[3] Neary's model extends that of Kemp, Yoshio Kimura, and Koji Okuguchi by allowing for permanent factor-market distortions. Kemp, Kimura, and Okuguchi demonstrated that their system is globally stable and that, following disturbances of specified kinds, factor allocations move in monotone fashion to their new stationary values. See also Kemp and Kimura.

[4] From (1),

$$\sigma=(\sigma_D-\sigma_X-\sigma_Y)|\theta|\cdot|\lambda|+\sigma_X R_X+\sigma_Y R_Y$$

where $R_i:=|\theta|\cdot|\lambda|+Q_i(i=X,Y)$.

Now

$$R_X=(\lambda_{LX}-\lambda_{KX})(\theta_{LX}-\theta_{LY})+\lambda_{LX}\theta_{KX}+\lambda_{Kx}\theta_{LX}$$

$$=\lambda_{LX}(\theta_{LX}+\theta_{KX})-\lambda_{LX}\theta_{LY}+\lambda_{KX}\theta_{LY}$$

$$=\lambda_{LX}\theta_{KY}+\lambda_{KX}\theta_{LY}>0$$

Similarly, $R_Y=\lambda_{KY}\theta_{LX}+\lambda_{LY}\theta_{KX}>0$ Hence for $\sigma>0$ it suffices that (2a) or (2b) be satisfied.

our Lemma 5 it is then easy to choose conditions which suffice for local stability and which are consistent with Neary's sufficient condition for fixed-price instability, viz., $|\theta|\cdot|\lambda|<0$. One such set of sufficient conditions consists of (d) of our Lemma 2 together with (b) and (c) or our Lemma 5.

Finally, one can extend Neary's original fixed-price model by allowing for accumulation while retaining the fixed-price assumption itself. This involves the reworking of Part II of our 1972 paper under the assumption of fixed commodity prices. Here we simply report that the fixed-price version of our earlier model also may be stable even if $|\theta|\cdot|\lambda|<0$.

REFERENCES

H. Herberg and M. C. Kemp, "Some Implications of Variable Returns to Scale," *Can. J. Econ.*, Aug. 1969, *2*, 403–15.

______ **and** ______, "Growth and Factor Market 'Imperfections'," *Z. ges. Staatswiss*. Nov. 1972, *128*, 590–604.

______, ______, **and S. P. Magee,** "Factor Market Distortions, the Shape of the Locus of Competitive Outputs, and the Relation between Product Prices and Equilibrium Outputs," *Econ. Rec.*, Dec. 1971, *47*, 518–30.

R. W. Jones, "Variable Returns to Scale in General Equilibrium Theory," *Int. Econ. Rev.*, Oct. 1968, *10*, 261–72.

______, "Distortions in Factor Markets and the General Equilibrium Model of Production," *J. Polit. Econ.*, June 1971, *79*, 437–59.

Murray C. Kemp, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

______ **and Yoshio Kimura,** *An Introduction to Mathematical Economics*, New York 1978.

______, ______, **and K. Okuguchi,** "Monotonicity Properties of a Dynamical Version of the Heckscher-Ohlin Model of Production," *Econ. Stud. Quart.*, Dec. 1977, *28*, 249–53.

W. Mayer, "Variable Returns to Scale in General Equilibrium Theory: A Comment," *Int. Econ. Rev.*, Feb. 1974, *15*, 225–35.

J. P. Neary, "Dynamic Stability and the Theory of Factor-Market Distortions," *Amer. Econ. Rev.*, Sept. 1978, *68*, 671–82.

This Side of Paradox, or, in Defense of the Correspondence Principle: A Reply to Herberg and Kemp

By J. PETER NEARY*

I am grateful to Horst Herberg and Murray Kemp (henceforth H-K) for their courteous response to my paper, but I regret to say that I disagree with their defense of the "paradoxes" in international trade theory discussed by Wolfgang Mayer and myself. Section I below reviews in detail the reasoning in my paper and in H-K's response, but the essential points in this exchange of views may, I believe, be summarized as follows.

In my paper it was demonstrated that, under a class of economically plausible assumptions about dynamic adjustment, a number of comparative static paradoxes which have attracted attention in the literature on factor-market distortions are associated with *unstable* equilibria. This association was shown to hold in both the two-sector Heckscher-Ohlin model of a small, open economy producing two traded goods, and the two-sector model with commodity prices determined endogenously. I concluded, by a standard application of Samuelson's Correspondence Principle, that these paradoxes can effectively be ignored since they will almost never be observed in either model.

Herberg and Kemp do not dispute my analysis but they question my conclusion on the grounds that stability of the larger static system (i.e., the endogenous-prices model) need not imply stability of the smaller one (i.e., the fixed-prices model) and that it is only the larger system which we should expect to be stable. Herberg and Kemp are certainly correct in pointing out that the stability conditions for the two models are different, but in attaching primacy to stability of one model rather than the other, they effectively call into question most applications of the Correspondence Principle. The essence of this principle is to make explicit the process by which a particular comparative static result does (or, in this case, does not) come about, on the grounds that if such a result is to be of economic, as opposed to purely mathematical, interest, it must refer to a specific temporal or causal sequence: for example, an increase in a subsidy will induce or give rise to (not merely "be associated with") a rise in output. Hence the only stability arguments which are relevant to comparative static statements made about the fixed-prices model are those which rely on dynamic adjustment mechanisms appended to *that* model; and similarly, the stability condition derived from the endogenous-prices model is relevant only to comparative static statements made about *that* model. Provided one adopts this viewpoint, which appears to be widely accepted in other branches of economic theory, it follows that H-K's arguments do not in any way invalidate the points made in my paper, since they do not rehabilitate the comparative static paradoxes I criticized.

I

Herberg and Kemp's main criticism of my paper is that the fixed-prices model of Sections I and II is "partial equilibrium in the sense that both commodity prices and factor endowments are taken to be unexplained constants," and that, if these parameters are made endogenous in a more general model, then this more general model may be stable when my model is unstable.

---

*Trinity College, Dublin. Without implicating them in the final product, I would like to thank Jagdish Bhagwati, Ronald Jones, Wolfgang Mayer and Frances Ruane for helpful comments, and the Institute for International Economic Studies, University of Stockholm, for providing research facilities.

In order to demonstrate that these statements are both correct and yet fully consistent with my paper, it is necessary to review the arguments in both papers. I first distinguish between the two models under consideration:[1]

Model α: The standard two-sector Heckscher-Ohlin model of international trade theory, with proportional intersectoral factor-price differentials, and with relative output prices parametrically fixed.

Model β: As model α, except that relative output prices are determined endogenously by the interaction of supply and demand, where aggregate demands are generated by the maximization of a homothetic utility function.

Consider next the following three statements:

A. Every observed equilibrium must be (locally) dynamically stable, with respect to a particular disequilibrium-adjustment mechanism.

B. The rankings of sectors by physical and value factor intensities are the same: $|\lambda||\theta|>0$.

C. The economy's distortion-output response is normal: an increase in the level of a permanent proportional factor-price differential imposed on a factor employed in one sector will, *ceteris paribus*, lower the output of that sector relative to that of the other sector.[2]

The main conclusion of Sections I and II of my paper was[3]

(1) $\alpha: A \leftrightarrow B$

while it is well-known in the literature (see the references cited in my original paper) that

(2) $\alpha: B \leftrightarrow C$

Combining (1) and (2), I concluded that

(3) $\alpha: A \leftrightarrow C$

In words, (3) states that when output prices are fixed, the hypothesis of dynamic stability rules out the comparative static paradox of a perverse distortion-output response. This chain of reasoning is of course a standard application of Samuelson's Correspondence Principle.

Having noted these results, H-K proceed to consider the endogenous-prices model β. They first quote the result of Section III of my paper that

(4) $\beta: A \leftrightarrow D$

where

D. The economy's aggregate elasticity of substitution σ is positive.

Then they show by algebraic manipulation that

(5) $\sim(D \leftrightarrow B)$

and they conclude from (4) and (5) that, by contrast with (1),

(6) $\beta: \sim(A \leftrightarrow B)$

or, in words: dynamic stability is consistent with a difference between the physical and value factor-intensity rankings of the two sectors, when output prices are endogenous.

Herberg and Kemp seem to imply that these conclusions are in conflict with the assertions in my paper. However, this is not the case. Statement (6) is explicitly stated in Section V of my paper (p. 679, point 2) and statement (5) clearly holds at point G in my Figure 3: condition B is not met there (since the supply curve in relative price-relative output space is downward sloping) but condition D is met (since the supply curve cuts the demand curve from below). Moreover it is noted in the text of my paper that at a stable equilibrium the distortion-output response of the endogenous-prices model is normal; i.e.,

---

[1] A more complete discussion of the models' assumptions is given in my paper. Note that model α is in fact a special case of model β, being the limiting case of the latter as the elasticity of substitution in demand tends towards infinity. As a result, the stability condition for model α, condition *B* below, is a special case of that for model β, condition *D* (provided the limit is taken in the correct manner; see Section III of my paper). However, in the present context it is more convenient to treat these models as distinct from each other.

[2] The discussion is couched in terms of the economy's distortion-output response, but it could alternatively have referred to the economy's price-output response (where the latter refers, when commodity prices are endogenous, to the response of aggregate supply to an exogenous upward shift in the demand schedule for one of the commodities). From a formal point of view, these two questions are essentially identical, since what is at issue in both cases is the response of relative outputs to a change in the prices faced by producers.

[3] Following standard convention, I use "&" for "and," "↔" for "implies and is implied by" and "∼" for "not." Note that all but one of the statements (1)–(8) are contingent on one or other of models α and β.

$$(7) \qquad \beta: D \leftrightarrow C$$

and so, from (4) and (7),

$$(8) \qquad \beta: A \leftrightarrow C$$

It was with the similarity between statements (3) and (8), and not the contrast between statements (1) and (6), that my paper was primarily concerned, and statement (6) does not invalidate the conclusions of my paper: a perverse price-output response is inconsistent with dynamic stability in both the fixed-prices model (from (3)) and the endogenous-prices model (from (8)).

The point of all this is that it is not helpful to ask whether or not a particular *stability condition* appropriate to one model is satisfied in a different model: conditions B and D are of relatively little intrinsic interest,[4] and even only slightly different static models imply different stability conditions.[5] Rather what matters is the information which the *hypothesis of stability* itself provides about the comparative static properties of each model, and in neither of the more general models considered by H-K have they shown that a comparative static paradox such as a nonfulfillment of property C is consistent with the stability hypothesis A.

II

Herberg and Kemp go beyond their specific criticisms of my paper to enunciate a general methodological principle when they state: "There seems to be no good reason to insist that any particular part of a larger dynamical system should be stable in isolation" (p. 812). While this argument is beyond reproach if the sole focus of analysis is the larger system, it is questionable in the present context on two counts. First, if a particular "small" model *should* be embedded in a larger one, then *neither* comparative static nor dynamic arguments are valid in isolation for the former. Whereas, if a comparative static statement is made about a particular small model, it is implicitly assumed that the economy under consideration will in fact move from the initial to the new equilibrium. The Correspondence Principle merely makes this movement explicit, and asserts that if the new equilibrium will not be approached, then the comparative static statement in question is of little economic interest. It follows that if this procedure is to make sense, it is essential that the dimensions of the comparative static and dynamic analyses be the same.

Second, even if we accept H-K's argument, where should we stop? Every conceivable model is in principle embedded in *some* more general model (except for that ultimate model which is a copy of, rather than an abstraction from, reality). Herberg and Kemp's larger statical system which allows for endogenous capital accumulation fails to allow for long-term changes in the rate of population growth or even longer-term changes in social institutions (such as the transition from feudalism to capitalism). It would clearly be farfetched to insist (as H-K must if they carry their general principle to its logical conclusion) that in searching for stability conditions to impose on comparative static calculations within the smaller system, one must look at a dynamic extension not of the smaller but of the largest conceivable static system. Herberg and Kemp's basic point would appear to be a recipe for infinite regress rather than for scientific progress in economic theory.

---

[4] I say "relatively" because there are some substantive issues to which conditions B and D are relevant in themselves. For example, in empirical tests of the Heckscher-Ohlin theorem, the question of whether or not condition B is met is relevant to the choice of data; see Stephen Magee, pp. 69–72.

[5] Both the increasing-returns model studied by Mayer and the two-sector fixed-prices model with a general (i.e., not necessarily proportional) intersectoral factor-price differential imply stability conditions different from B.

III

It should be clear from the foregoing that H-K and I do not disagree on the specification or properties of these models but only on questions of their interpretation. We are also in agreement on a further point: what I called the Ambiguity of the Correspondence Principle, by which I meant the well-known point that the stability conditions implied by one set of assumptions about disequilibrium adjustment, may not apply under a different set. However, while H-K suggest that because of this the conclusions of my paper are "highly tentative," it is my belief that they are likely to prove highly robust.[6] For example, as I noted in the concluding section of my paper, the stability condition B continues to apply under more general labor market adjustment assumptions than those implicit in equation (14) of my paper.[7]

Indeed, I have been able to find only one class of adjustment mechanisms under which the fixed-prices model with proportional intersectoral factor-price differentials is *unstable* when condition B holds. A typical member of this class is given by equations (13) and (14) of my paper, with ϕ' *negative* instead of positive. This is equivalent to assuming that adjustment in the capital market depends on the degree of disequilibrium in the *labor* market (i.e., capital flows into the sector where the lower rental and hence the higher wage is being earned), and it is therefore a simple application of the argument with which Don Patinkin (see his Appendix to ch. 7) attacked the generality of the Correspondence Principle. But it is clearly a highly implausible assumption, and the fact that it gives rise to paradoxical comparative static predictions reflects not the Correspondence Principle, but the Principle of the Conservation of (Waste) Matter (better known as "garbage in, garbage out"). Until H-K or anyone else produce an economically plausible adjustment mechanism which implies stability conditions for these models which differ from B and D, I will continue to argue that paradoxes such as the nonfulfillment of statement C are theoretical curiosa that will almost never be observed.

REFERENCES

H. Herberg and M. C. Kemp, "In Defense of Some 'Paradoxes' of Trade Theory," *Amer. Econ. Rev.*, Sept 1980, *70*, 812–14.

Stephen P. Magee, *International Trade and Distortions in Factor Markets*, New York 1976.

W. Mayer, "Variable Returns to Scale in General Equilibrium Theory: A Comment," *Int. Econ. Rev.*, Feb. 1974, *15*, 225–35.

J. P. Neary, "Dynamic Stability and the Theory of Factor-Market Distortions," *Amer. Econ. Rev.*, Sept. 1978, *68*, 671–82.

———, "Wage Stickiness, Inter-sectoral Capital Reallocation, and the Case for Adjustment Assistance," in Jadish N. Bhagwati and T. N. Srinivasan, eds., *Import Competition and Adjustment Policy*, Chicago forthcoming.

Don Patinkin, *Money, Interest and Prices*, 2d ed., New York 1965.

---

[6] Note that the type of robustness at issue here is different from that considered in Section I above. There it was noted that the stability condition B is *not* robust with respect to changes in the static model (i.e., it applies only to the fixed-prices two-sector Heckscher-Ohlin model α), whereas here it is argued that this condition *is* robust with respect to changes in the dynamic adjustment assumptions which are appended to that particular static model.

[7] See my forthcoming paper.

Utility Function Transformations and Money Illusion: Comment

By PETER HOWITT AND DON PATINKIN*

In a recent exchange in this *Review*,[1] Richard Dusansky and Peter Kalman (hereafter, D-K) have claimed that the standard condition that utility functions be homogeneous of degree zero (*HD0*) in money and prices is not a necessary one for demand functions to be free of money illusion. They have proposed a "more general" set of conditions that imply illusion-free demand functions. Clower and Riley (hereafter, C-R) have claimed that, on the contrary, the standard condition serves completely to characterize the class of illusion-free demand functions, and that the conditions of D-K achieve only a spurious generality. However, D-K have apparently rebutted this claim by pointing out what is actually a minor error in the paper by C-R.[2]

The purpose of the present note is to confirm C-R's claim by demonstrating the nature of D-K's spurious generality in a simple way that goes back to fundamental utility theory. In particular, we shall show that D-K's conditions are more general only in the sense that they allow a consumer whose ordinal utility is *HD0* to choose an index of cardinal utility that depends upon the vector of money prices. The spuriousness of this generality arises from the well-known fact that (under conditions of certainty) no operational significance can be attached to the choice of a cardinal utility index. Correspondingly, D-K's conditions are operationally equivalent to the standard *HD0* condition, in the sense that they generate the same demand functions.

More specifically, consider one consumer with the utility function $u(x)$ and a second one with the utility function $f(x)$ defined by

$$f(x) \equiv \phi(u(x)) \tag{1}$$

where $x=(x_1,\ldots,x_n)$ is the consumption vector, and ϕ is a strictly increasing function, which (following Paul Samuelson) we will call the consumer's index of cardinal utility. It is a commonplace of consumer theory[3] that the marginal rates of substitution of two such consumers are identical, for

$$\frac{\dfrac{\partial f}{\partial x_i}(\;)}{\dfrac{\partial f}{\partial x_j}(\;)} \equiv \frac{\dfrac{d\phi}{du}(\;)\cdot\dfrac{\partial u}{\partial x_i}(\;)}{\dfrac{d\phi}{du}(\;)\cdot\dfrac{\partial u}{\partial x_j}(\;)} \equiv \frac{\dfrac{\partial u}{\partial x_i}(\;)}{\dfrac{\partial u}{\partial x_j}(\;)} \tag{2}$$

Correspondingly, each consumer will exhibit exactly the same demand functions. Thus the choice of the cardinal index has no operational significance.

Assume now that the second consumer chooses his cardinal index in accordance with the phases of the moon; that is, when the phase of the moon is λ he chooses the cardinal index $\phi^{\lambda}(\;)$. In this case the consumer's utility function can be defined as

$$f(x,\lambda) \equiv \phi^{\lambda}(u(x)) \tag{3}$$

One might get the impression that a consumer with such a utility function could be called a lunatic in the primal sense of the term, because his utility is affected by lunar phases. However, it would be misleading to claim that we had in this way generalized the standard theory of consumer's behavior to incorporate the case of lunacy. For in the present case the only role played by λ is to choose the inconsequential cardinal index which cancels out as in (2) above. In other words, the utility function defined by (3) deals only with lunatics who say they feel better (or worse) when the full moon comes out, but do not behave any differently. Thus it allows only a spurious generality.

---

*Associate professor of economics, University of Western Ontario, and professor of economics, Hebrew University of Jerusalem, respectively. Howitt acknowledges financial support from the Canada Council. Work on this note was begun during Patinkin's visit at the University of Western Ontario in the fall of 1976. His subsequent work was carried out as a fellow of the Institute for Advanced Studies at the Hebrew University; it was financed in part by a grant from the Ford Foundation, administered by the Maurice Falk Institute for Economic Research in Israel. We are greatly indebted to Menahem Yaari of the Hebrew University for most helpful discussions which took place while he too was a fellow of the Institute for Advanced Studies.

[1] See Dusansky and Kalman (1974, 1976) and Robert Clower and John Riley.

[2] The error was C-R's omission (equation (3), p. 184) of the inconsequential $h(p)$, as in our equation (8) below.

[3] See Vilfredo Pareto, pp. 539–45, John Hicks, pp. 306–07, and Paul Samuelson, pp. 98–99.

Along the same lines, we might suppose that instead of looking at phases of the moon, the second consumer looks at the vector of prices, $p=(p_1,\dots,p_n)$ when deciding which cardinal utility index to use. Thus the utility function would be defined as

$$f(x,p)\equiv\phi^p(u(x)) \tag{4}$$

where for each p, ϕ^p is monotonically increasing. Just as before, one would hardly want to claim that this involves a generalization of the standard theory to incorporate Veblenesque snob effects by virtue of the inclusion of the price vector in the utility function: for once again the marginal rates of substitution of the two consumers will be identical, in view of the fact that

$$\frac{\frac{\partial f}{\partial x_i}(\ )}{\frac{\partial f}{\partial x_j}(\ )}\equiv\frac{\frac{d\phi^p}{du}(\ )\cdot\frac{\partial u}{\partial x_i}(\ )}{\frac{d\phi^p}{du}(\ )\cdot\frac{\partial u}{\partial x_j}(\ )}\equiv\frac{\frac{\partial u}{\partial x_i}(\ )}{\frac{\partial u}{\partial x_j}(\ )} \tag{5}$$

In the standard theory of consumer choice with money, the consumer is assumed to possess a utility function $u(x,m,p)$ which is *HD0* in m and p, where m is the nominal demand for money. The consumer maximizes this utility function subject to the budget constraint

$$\sum_{i=1}^{n} p_i x_i+m=W \tag{6}$$

where W denotes the money value of the consumer's initial endowment. The budget constraint (6), together with the first-order conditions:

$$\frac{\partial u}{\partial x_i}(\ )/\frac{\partial u}{\partial m}(\ )=p_i \quad i=1,\dots,n \tag{7}$$

define the demand functions $\bar{x}(p,W)$, $\bar{m}(p,W)$. These functions are easily seen to be free of money illusion; i.e., $\bar{x}(\ )$ is *HD0* and $\bar{m}(\ )$ is homogeneous of degree one (*HD1*) in p and W.[4]

Just as before, the choice of the particular cardinal utility index does not affect the marginal rates of substitution in (7) and hence has no operational significance. In particular, the consumer's demand functions would be just the same if his utility function were defined by $u(x,m,p)$ or by $\phi^\lambda[u(x,m,p)]$ or by $\phi^p[u(x,m,p)]$. In this last case, the utility function generally would not be *HD0* in m and p. For if m and p were doubled, the value of $u(\ )$ would be unaffected; but since the change in p might dictate the choice of a different cardinal utility index, the value of $\phi^p[u(\ )]$ might be affected. However, it is obvious that such a utility function would not imply any operationally meaningful form of money illusion.

Now, this is really what is involved in D-K's so-called "generalization" of the standard theory. For despite the technical complexity of their contribution, the class of utility functions defined by their proposed conditions is simply

$$f(x,m,p)\equiv(p_1)^k u(x,m,p)+h(p) \tag{8}$$

where $u(\ )$ is differentiable and *HD0* in m and p, and $h(p)$ is any function. Their conditions (1974, p. 118) are that $\partial f(\ )/\partial x_i$ be homogeneous of degree k for all i, and that $\partial f(\ )/\partial m$ be homogeneous of degree $k-1$, in m and p.[5] Define $u(x,$

---

[4] To see this note that if $u(\ )$ is *HD0*, then the marginal rates of substitution in (7) are *HD1* in m and p (see R. G. D. Allen, p. 324). Thus if m is doubled but x unchanged when p and W are doubled, (6) and (7) will continue to hold. See also Patinkin, pp. 289–90.

[5] Actually D-K would have seen immediately that something was wrong with these conditions if they had noticed that the conditions apply directly to marginal utility functions, which in ordinal utility theory have no operational significance except indirectly through their ratios, the marginal rates of substitution.

$m, p) \equiv f(x, m/p_1, p/p_1)$ and $\hat{h}(x, m, p) \equiv f(x, m, p) - (p_1)^k u(x, m, p)$. Then (8) holds if $\hat{h}(x, m, p)$ can be expressed as a function only of p. This will be the case if $\partial \hat{h}(x, m, p)/\partial x_i \equiv \partial \hat{h}(x, m, p)/\partial m \equiv 0$ for all i; and this condition follows immediately from D-K's homogeneity conditions. That (8) implies D-K's conditions follows immediately from the fact that, because $u(\ )$ is *HD0*, therefore $\partial u(\ )/\partial x_i$ is *HD0* for all i, and $\partial u/\partial m$ is homogeneous of degree minus one, in m and p (see Allen, p. 324).

Clearly every utility function in this class is of the form $\phi^p[u(x, m, p)]$, where, for any given p, $\phi^p[\ ]$ is an increasing function. Thus the only sort of added generality allowed for by D-K's conditions is the spurious generality of permitting the consumer to vary the cardinal utility index when prices are changed. As C-R put it, the only sense in which the standard *HD0* condition is, as D-K claim, overly restrictive, is that it ignores the case in which the consumer says he feels better (or worse) when money holdings and all money prices are doubled, but does not behave any differently.

To set the record straight, we show that a necessary and sufficient condition for demand functions to be free from money illusion is that they be derivable from a utility function that is *HD0* in m and p. Sufficiency is proven in footnote 4. To prove necessity, suppose that the illusion-free demand functions $\bar{x}(\ )$, $\bar{m}(\ )$ are derivable from the utility function: $f(x, m, p)$, which is not *HD0*. Then it suffices to show that they are also derivable from the utility function: $u(x, m, p) \equiv f(x, m/p_1, p/p_1)$, which is *HD0*. Take any values (p^0, W^0) of prices and wealth, and define $x^0 \equiv \bar{x}(p^0, W^0)$, $m^0 \equiv \bar{m}(p^0, W^0)$. We want to show that x^0 and m^0 are also the quantities that will be chosen by the consumer with utility function $u(\ )$ when faced with (p^0, W^0). This will be the case if the first-order conditions

$$(9) \qquad \frac{\dfrac{\partial u}{\partial x_i}(x^0, m^0, p^0)}{\dfrac{\partial u}{\partial m}(x^0, m^0, p^0)} = p_i^0 \qquad i = 1, \ldots, n$$

are satisfied. To show that they are indeed satisfied, note that, since $\bar{x}(\ )$ and $\bar{m}(\ )$ are free of money illusion, therefore $x^0 = \bar{x}(p', W')$ and $m' = \bar{m}(p', W')$; where $(p', W', m') \equiv (1/p_1^0)(p^0, W^0, m^0)$. Therefore, since $\bar{x}(\ )$ and $\bar{m}(\ )$ are derivable from $f(\ )$, the first-order conditions:

$$(10) \qquad \frac{\dfrac{\partial f}{\partial x_i}(x^0, m', p')}{\dfrac{\partial f}{\partial m}(x^0, m', p')} = p_i' \qquad i = 1, \ldots, n$$

must hold. Equation (9) can be derived directly from (10) using the chain rule of calculus.

Thus D-K have merely shown that illusion-free demand functions may also be derivable from operationally equivalent cardinal utility functions that are not *HD0*. Indeed D-K could have gone further and noted that they may be derivable from any utility function that has the same marginal rates of substitution as an *HD0* function whenever (7) is satisfied;[6] but once again this added generality has no operational significance. From the operational viewpoint, then, the traditional theory of money illusion is reaffirmed.

---

[6] When (7) is not satisfied the utility functions need not satisfy any homogeneity conditions, because these values of (x, m, p) can never correspond to a point on the demand function. See Robert Pollak, pp. 70–71. This set of utility functions is not of the form $\phi^p[u(x, m, p)]$ where $u(\ )$ is *HD0*. Nor, as Menahem Yaari has pointed out to us, is the transformation employed in our proof that any illusion-free demand function is derivable from an *HD0* utility function. This means that the class of utility functions that are operationally equivalent to *HD0* functions is broader than the class of monotonic transformations, which in turn is broader than the class defined by D-K's conditions.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, New York 1938.

R. W. Clower and J. G. Riley, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model: Comment," *Amer. Econ. Rev.*, Mar. 1976, *66*, 184–85.

R. Dusansky and P. J. Kalman, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model," *Amer. Econ. Rev.*, Mar. 1974, *64*, 115–22.

______ **and** ______, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model: Reply," *Amer. Econ. Rev.*, Mar. 1976, *66*, 192–95.

John R. Hicks, *Value and Capital*, Oxford 1939.

Vilfredo Pareto, *Manuel d'économie politique*, Paris 1909.

Don Patinkin, *Money, Interest, and Prices*, Evanston 1956.

R. A. Pollak, "Price Dependent Preferences," *Amer. Econ. Rev.*, Mar. 1977, *67*, 64–75.

Paul A. Samuelson, *Foundations of Economic Analysis*, Cambridge Mass. 1947.

Utility Function Transformations and Money Illusion: Reply and Further Results

By RICHARD DUSANSKY*

In our 1974 paper, Peter Kalman and I studied the theory of consumer choice with money, in which a generalized real balance effect enters the utility function. It was shown that the class of utility functions which yields illusion-free commodity demand functions (*IFDF*) contains utility functions which are nonhomogeneous. Peter Howitt and Don Patinkin assert that the generalization of the class beyond the "standard" (i.e., homogeneous of degree zero, *HD0*) utility function is spurious, in that the nonhomogeneous functions in the *IFDF* class do not differ *ordinally* from the standard utility function.

Provided below is a complete mathematical characterization of the entire *IFDF* utility class (in the form of an "if and only if" proof). It is shown that the *IFDF* class contains nonhomogeneous utility functions which differ in an ordinal way from the standard utility function.

Consider the "regular" utility function

$$u=f(x,m,p) \tag{1}$$

where x represents the $n-1$ commodities, $x_1,\dots,x_{n-1}$, m is holdings of nominal money balances, and p represents the $n-1$ commodity prices, $p_1,\dots,p_{n-1}$.

The budget constraint is

$$y+L=\sum_{i=1}^{n-1} p_i x_i+m \tag{2}$$

where y is income and L is initial money balances. (Note that the price of m is taken to be unity.) The maximization of (1) over x, m subject to (2) yields the first-order conditions

$$MRS_{ij}(x,m,p)=\frac{\dfrac{\partial f(x,m,p)}{\partial x_i}}{\dfrac{\partial f(x,m,p)}{\partial x_j}}=\frac{p_i}{p_j} \tag{3}$$

$$i=1,\dots,j-1,j+1,\dots,n-1$$

$$MRS_{mj}(x,m,p)=\frac{\dfrac{\partial f(x,m,p)}{\partial m}}{\dfrac{\partial f(x,m,p)}{\partial x_j}}=\frac{1}{p_j}$$

$$y+L=\sum_{i=1}^{n-1} p_i x_i+m$$

which, under the usual conditions, can be solved for the demand functions

$$x_i=h^i(p,L,y) \qquad i=1,\dots,n-1 \tag{4}$$

$$m=h^m(p,L,y)$$

These demand functions, which exist in the neighborhood of the solution, are unique. In general these demand functions exhibit money illusion.

Now, the central concern. What is the necessary and sufficient condition for the commodity demand functions in (4) to be free from money illusion? The following proposition is put forth and proved.

PROPOSITION: *The necessary and sufficient condition under which the commodity demand functions are HD0 in* (p,L,y) *is*

$$\sum_{j,k=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial x_k\partial p_j}\frac{D_{ki}}{D}p_j \tag{5}$$

$$+\sum_{j=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial m\partial p_j}\frac{D_{ni}}{D}p_j$$

$$+\lambda\frac{D_{ni}}{D}+m\frac{D_{n+1i}}{D}=0$$

---

*Professor of economics, State University of New York-Stony Brook.

PROOF:

By Euler's theorem on homogeneous functions we know

$$(6)\qquad h^i(p,L,y)=h^i(\alpha p,\alpha L,\alpha y)$$

if and only if

$$(7)\qquad \sum_{j=1}^{n-1}\frac{\partial h^i}{\partial p_j}p_j+\frac{\partial h^i}{\partial L}L+\frac{\partial h^i}{\partial y}y=0 \qquad i=1,\dots,n-1$$

However, from a total differentiation of the first-order conditions we know that[1]

$$\frac{\partial h^i}{\partial p_j}=\lambda\frac{D_{ji}}{D}-\sum_{k=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial x_k\partial p_j}\frac{D_{ki}}{D}-\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial m\partial p_j}\frac{D_{ni}}{D}+x_j\frac{D_{n+1i}}{D}$$

$$\frac{\partial h^i}{\partial y}=\frac{\partial h^i}{\partial L}=-\frac{D_{n+1i}}{D}$$

where D is the determinant of the Jacobian of the first-order conditions, D_{ji} is the cofactor of the element in row j column i of the Jacobian, and λ is the Lagrange multiplier. Hence, we have (7) if and only if

$$(8)\qquad \sum_{j=1}^{n-1}\left[\lambda\frac{D_{ji}}{D}-\sum_{k=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial x_k\partial p_j}\frac{D_{ki}}{D}-\frac{\partial^2(\cdot,\cdot,\cdot)}{\partial m\partial p_j}\frac{D_{ni}}{D}+x_j\frac{D_{n+1i}}{D}\right]p_j-[L+y]\frac{D_{n+1i}}{D}=0$$

Using (2) we have (8) if and only if

$$(9)\qquad \frac{\lambda}{D}\sum_{j=1}^{n-1}D_{ji}p_j-\sum_{j=1}^{n-1}\sum_{k=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial x_k\partial p_j}\frac{D_{ki}}{D}p_j-\sum_{j=1}^{n-1}\frac{\partial^2 f(\cdot,\cdot,\cdot)}{\partial m\partial p_j}\frac{D_{ni}}{D}p_j-m\frac{D_{n+1i}}{D}=0$$

However, by the theorem on expansion by alien cofactors, we know that

$$\sum_{j=1}^{n-1}D_{ji}p_j+D_{ni}=0, \qquad i=1,\dots,n-1$$

Hence, (9) if and only if (5).

The necessary and sufficient condition in (5) constitutes the most general condition under which the commodity demand functions are *HD0*, and completely characterizes the class of utility functions which yields illusion-free commodity demand functions (i.e., the *IFDF* class). It tells us that the economic responses to the equiproportionate change in p, L, y (i.e., the money illusion experiment), as it impacts through the marginal utility functions and through the budget constraint, must of course be completely offsetting in equilibrium, so that commodity demands remain unchanged. Much more important, the condition in (5) has useful implications. Note that (5) imposes no homogeneity requirement of any kind on the utility function, on the marginal utility functions, or on the marginal rate of substitution functions. The last is particularly important. It is possible for the indifference surfaces to shift in response to the money illusion experiment, and for the marginal rates of substitution in (3) to vary, and still for demand to remain unaltered. We are reminded that the *MRS*s are *functions*, having x, m, and p as arguments, which (for the illusion-free demand outcome) need to exhibit unchanged commodity demand values at the solution point (i.e., in equilibrium), but need not themselves be invariant. The complete characterization of the *IFDF* class thus admits utility functions whose commodity *MRS* functions are nonhomogeneous.[2]

---

[1] For a detailed derivation, see my paper with Kalman (1972, pp. 343-45).

[2] It may be the case that standard utility functions generate all illusion-free demand functions. Just as it may be true that nonhomogeneous utility functions also generate all illusion-free demand functions. The direct characterization of demand, however, is not the issue. The issue is characterization of utility, i.e., the class of utility functions which yield illusion-free demand functions.

Now consider the class of utility functions which are derivable from a standard (i.e., *HD0*) utility function by order-preserving transformations. Let $u(x,m,p)$ be *HD0* in m and p, and ϕ be any positive monotonic transformation of $u(x,m,p)$. The resulting class is given by

$$\tilde{u}(x,m,p)=\phi[u(x,m,p)],\ \phi'[u]>0 \tag{10}$$

It is easy to show, since $u(x,m,p)$ is *HD0* in m and p, that *all* utility functions in this class exhibit *homogeneous MRS* functions. The class obviously does not contain utility functions which exhibit *non*homogeneous *MRS* functions. However, we know from the complete characterization of the *IFDF* class derived above that there exist some of the latter (i.e., some nonhomogeneous utility functions having nonhomogeneous *MRS* functions) which yield illusion-free demand functions. Hence, since these nonhomogeneous utility functions are not contained in (10) it must be the case that they differ ordinally from the standard utility function.

A by-product of the complete characterization of the *IFDF* class is a release from the use of homogeneity restrictions as a means for attaining illusion-free demand functions. The results free monetary economists from fixation on homogeneity (in the utility function, in the marginal utility functions, or in the *MRS* functions) and on the invariance of the *MRS* functions. A homogeneity assumption is no longer required in order to have a homogeneity result. That these nonhomogeneous utility functions exist is important in completely characterizing utility theory.[3]

---

[3] The results reported herein constitute a further, and final, development of the analysis of money illusion first initiated in my 1974 paper with Kalman, and should be taken as representing our final position. A more complete exposition appears in my 1977 paper with Kalman.

REFERENCES

R. W. Clower and J. G. Riley, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model: Comment," *Amer. Econ. Rev.*, Mar. 1976, *66*, 184–85.

R. Dusansky and P. J. Kalman, "The Real Balance Effect and the Traditional Theory of Consumer Behavior: A Reconciliation," *J. Econ. Theory*, Dec. 1972, *5*, 336–47; "Erratum," Feb. 1973, *6*, 107.

______ **and** ______, "The Foundation of Money Illusion in a Neoclassical Micro-Monetary Model," *Amer. Econ. Rev.*, Mar. 1974, *64*, 115–22.

______ **and** ______, "The Foundation of Money Illusion in a Neoclassical Micro-Monetary Model: Reply," *Amer. Econ. Rev.*, Mar. 1976, *66*, 192–95.

______ **and** ______, "Illusion-Free Demand in a Monetary Economy," mimeo., State Univ. New York-Stony Brook 1977.

P. Howitt and D. Patinkin, "Utility-Function Transformations and Money Illusion: Comment," *Amer. Econ. Rev.*, Sept. 1980, *70*, 819–22.

Utility Function Transformations and Money Illusion: A Further Comment

By PETER HOWITT AND DON PATINKIN*

Our preceding comment in this issue makes two points:

1) A specific one, that the Dusansky-Kalman (D-K) papers (1974, 1976) specify utility functions which are actually monotonic transformations of standard ones (i.e., functions which are homogeneous of degree zero (*HD0*) and thus provide only a spurious generalization of the class of utility functions that generate demand functions which are free of money illusion.

2) A general one, consisting of the demonstration of the basic proposition that a necessary and sufficient condition for demand functions to be free of money illusion is that they be derivable from a standard utility function.

Ostensibly, Dusansky has provided a "reply" to our comment; actually he has not dealt with either of the preceding points. Thus, instead of defending the original D-K conditions on utility functions, he implicitly accepts our criticism of them by replacing them with yet another set of conditions. He then attempts to create the impression that this new set of conditions somehow refutes our foregoing basic proposition; actually, however, the proposition which Dusansky attributes to us and devotes his reply to refuting with the aid of some elaborate mathematics is one that we have explicitly pointed out to the reader that we have *not* made.

Specifically, Dusansky attributes to us the view that nonhomogeneous utility functions which yield illusion-free demand functions "do not differ *ordinally* from the standard utility function" (p.823). What our foregoing basic proposition actually claims and proves, however (and Dusansky has not challenged the validity of this proof), is that they do not differ *operationally* from the standard utility function, that is, they yield the same demand functions; and we explicitly emphasize in footnote 6 of our comment that our proof implies that "the class of utility functions that are operationally equivalent to *H0)* functions is broader than the class of monotonic [i.e., order-preserving] transformations" (p.821). Because of this failure on Dusansky's part to distinguish between "operationally" and "ordinally," he also fails to realize that his new conditions, though free of the specific defect that we noted in the original ones, nevertheless share with them the general defect that they provide no more than a spurious generalization of the standard class of utility functions.[1]

The basic point here (and the one essentially made in our comment, fn. 6) is that once prices enter utility functions, it is no longer true, as it is in standard consumer theory, that for two utility functions to yield the same demand function they must be positive monotonic transformations of each other. The intuitive explanation of this fact is that the entry of prices into the utility function creates an additional interdependence between it and the budget restraint that precludes certain combinations of (x, m, p) from ever being chosen by the household as optimal ones. Hence two different utility functions—$u(x, m, p)$ and $f(x, m, p)$—need not provide the same ordering of these "unchooseable" combinations in order to yield the same demand functions.

As a simple illustration of this point, consider the case of a single consumable good, where $n=1$, so that budget restraint (6) of our note reduces to

---

*Associate professor of economics, University of Western Ontario, and professor of economics, Hebrew University of Jerusalem, respectively.

[1] In fn. 2 of his reply, Dusansky effectively says that even nonoperational differences between utility functions constitute an issue for him. It is difficult to understand why.

$$x+\frac{m}{p}=\frac{W}{p}$$

It follows that every possible budget line in $(x, m/p)$ space has a slope of -1 (see Figure 1). Consider now two different utility functions which generate the same illusion-free demand functions, represented by the expansion path $0E$ in the figure (the absence of money illusion manifests itself in the uniqueness of the expansion path in $(x, m/p)$ space; i.e., in its dependence on the ratio of m to p, and not on their individual values). Let the first utility function be a nonstandard (i.e., nonhomogeneous) one, say $h(x, m, p)$, whose money illusion manifests itself in the fact that its indifference curves in the figure cannot be drawn without first specifying the price level p. (Thus all variations in m/p in the figure are now assumed to be variations in m with p held constant.) Suppose in particular that this utility function has the solid indifference curves when the price level is p_0, but has the dashed indifference curves when the price level is p_1. Since $h(x, m, p)$ is assumed to generate the expansion path $0E$, the points on both sets of these curves whose marginal rate of substitution of x for m/p is unity must all lie along $0E$; and the same would be true for the indifference curves of $h(x, m, p)$ drawn for any value of p.

Now, our basic proposition states that the illusion-free expansion path $0E$ can also be derived from a standard, illusion-free utility function, say $g(x, m/p)$. Clearly, the points on the indifference curves of this function whose marginal rate of substitution of x for m/p is unity must also all lie along $0E$. For simplicity, let these indifference curves also be represented by the solid ones in the figure. From the fact that these curves intersect with the dashed ones (at which points, of course, the marginal rates of substitution must differ) it is clear that the utility functions $g(x, m/p)$ and $h(x, m, p)$ do not represent the same system of ordinal preferences: that is, they are not positive monotonic transformations of one another. It is, however, also clear that this difference between the utility functions is not observable, and hence not operational, for their marginal rates of substitution can differ only at points that the consumer would in any event never choose.[2]

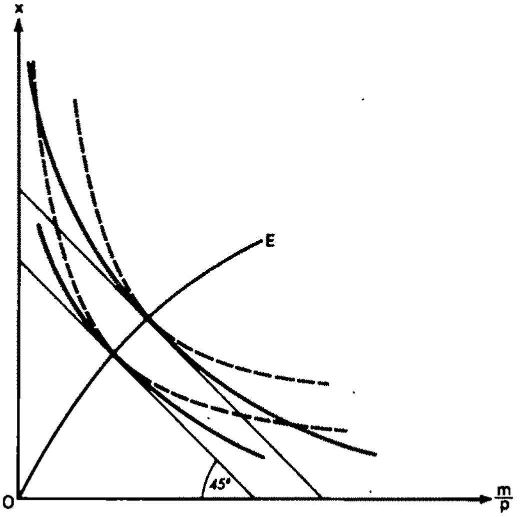


FIGURE 1

We note finally that this figure enables us to provide an even more meaningful statement of the necessary and sufficient condition for generating illusion-free demand functions. We begin by observing that the assumption that the nonhomogeneous utility function $h(x, m, p)$ generates the illusion-free expansion path OE implies that the marginal rate of substitution of x for m/p specified by this utility function for any point on this path depends only on the ratio of m to p, and not on their individual values. This means that even if the utility function is not *HD0* in m and p at these points, the function representing the marginal rate of substitution must be; conversely, the marginal rate of substitution function can be nonhomogeneous only at points that are off the expansion path, and hence nonobservable. Thus a necessary and sufficient condition for demand functions to be *HD0* in m and p is that the marginal rate of substitution function be *HD0* in these variables at all observable points.[3] This proposition is almost self-evident; we have stated it here only because of the necessity to set the record straight.[4]

---

[2] It has not escaped our notice that this figure suggests an obvious proof that, in contrast to standard consumer theory, once prices enter utility functions there are generally no integrability conditions allowing one to infer a unique ordinal preference mapping from observed demand behavior. See also Robert Pollak, p. 70.

REFERENCES

R. Dusansky, "Utility Function Transformations and Money Illusion: Reply," *Amer. Econ. Rev.*, Sept. 1980, *70*, 823–25.

______ **and P. J. Kalman,** "The Foundation of Money Illusion in a Neoclassical Micro-Monetary Model," *Amer. Econ. Rev.*, Mar. 1974, *64*, 115–22.

______ **and** ______, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model: Reply," *Amer. Econ. Rev.*, Mar. 1976, *66*, 192–95.

P. Howitt and D. Patinkin, "Utility Function Transformations and Money Illusion: Comment," *Amer. Econ. Rev.*, Sept. 1980, *70*, 819–22.

R. A. Pollak, "Price Dependent Preferences," *Amer. Econ. Rev.*, Mar. 1977, *67*, 64–75.

---

[3] In the case of $n>1$, there are the usual technical problems of defining an appropriate price index p by which to deflate nominal money balances. In order to avoid such problems, this condition should then be stated in the following equivalent manner: that the *MRS* functions $(\partial h/\partial x_i)/(\partial h/\partial m)$ $(i=1,\ldots,n)$ be homogeneous of degree one in m and p at all observable points. This can readily be demonstrated from utility-maximization conditions (6)–(7) of our original comment, with $u(\ )$ replaced by $h(\ )$.

[4] In particular, Dusansky is in error when he states that the necessary and sufficient condition for commodity demand functions to be *HD0* "imposes no homogeneity requirement of any kind on the utility function, on the marginal utility functions, or on the marginal rate of substitution functions. The last is particularly important" (p. 824). Dusansky's error stems from his failure to examine the properties of the marginal rate of substitution functions at the observable points—which are, of course, the only operationally meaningful ones.

On the Comparison of the Stability Implications of Marshallian and Walrasian Adjustment Schemes: Note

By IAN M. MCDONALD*

Axel Leijonhufvud has argued that

> *In the Keynesian macrosystem the Marshallian ranking of price- and quantity-adjustment speeds is reversed*: In the shortest period flow quantities are freely variable, but one or more prices are given, and the admissible range of variation for the rest of prices is thereby limited.... The "revolutionary" element in the *General Theory* can perhaps not be stated in simpler terms [p. 52].

This has become a highly influential view. For example, Milton Friedman concurred when he argued that Keynes "... deviated from Marshall, and it was a momentous deviation,... in reversing the roles assigned to price and quantity. He assumed that, at least for changes in aggregate demand, quantity was the variable that adjusted rapidly, while price was the variable that adjusted slowly, at least in a downward direction" (p. 209).

More recently James Tobin has argued that one way to appreciate the importance of this change of dynamic structure is to analyze the stability implications of Marshallian and Walrasian adjustment processes. Tobin argues persuasively that Keynes' achievement was to show that, in a capitalist economy, market forces may be too weak to generate a full-employment equilibrium. Describing Keynes' achievement in this way renders irrelevant the Pigou argument that Keynes failed to establish the existence of an underemployment equilibrium. Thus Tobin writes "[t]he real issue is not the existence of a long-run static *equilibrium* with unemployment, but the possibility of protracted unemployment which the natural adjustments of a market economy remedy very slowly if at all" (pp. 195–96).

In the light of these arguments Tobin develops two simple macro models, one reflecting the Walras-Keynes-Phillips (W-K-P) adjustment relation and the other reflecting the Marshallian (M) adjustment relation. It is clear from Tobin's analysis that the speed of adjustment of price and quantity is not the only characteristic distinguishing the Marshallian and Walrasian approaches. The functional specification of the two adjustment processes is also different in that "Marshallian adjustment in a particular market is that quantity adjusts to the difference between demand price and supply price for existing quantity. Walrasian adjustment is that quantity adjusts to the difference between demand and supply at existing price" (Tobin, p. 196).

This comment endeavors to make the following points on Tobin's analysis:

1) Tobin's statement of the stability condition of the W-K-P model is incorrect—the correct stability condition is in fact more restrictive.

2) Tobin's specification of the M model is, as Tobin himself is aware, "... separable into output and price equations." Furthermore the output equation "... is a stable differential equation ..." (p. 200). This avoids the question of whether the Marshallian adjustment mechanism will in fact yield a stable model in output as well as price. An output equation in keeping with Tobin's statement of the Marshallian adjustment process is specified below. The condition for stability is derived and shown to be weaker than the W-K-P condition.

---

*Lecturer in economics, University of Melbourne.

I

Tobin's W-K-P model consists of the following equations

$$(1)\qquad DY=A_y(E(P,x,Y)-Y)$$

$$(2)\qquad \frac{DP}{P}=\Pi=A_p(Y-Y^*)+x$$

$$(3)\qquad Dx=A_x(\Pi-x)$$

where Y = aggregate output
E = aggregate demand $=E(P,x,Y)$
P = price level
x = expected rate of inflation

and A_y, A_p, $A_x>0$ and the derivatives of the function of E are restricted by $E_p<0, E_x, E_y>0$.

By substituting (2) into (3) and linearizing the resulting equations, the system can be written as

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}=\begin{bmatrix} A_y(E_y-1)-D & A_yE_p & A_YE_x \\ A_pP^* & -D & P^* \\ A_xA_p & 0 & -D \end{bmatrix}\cdot\begin{bmatrix} Y-Y^* \\ P-P^* \\ x \end{bmatrix}$$

where $Y^*=Y$ and $P^*=P$ are the equilibrium values.

To derive the stability conditions, I derive the characteristic equation of the matrix. This can be written

$$0=a_0D^3+a_1D^2+a_2D+a_3$$

where $a_0=1$
$a_1=-A_y(E_y-1)$
$a_2=-A_yA_p(E_pP^*+E_xA_x)$
$a_3=-A_yA_pE_pP^*A_x$

The Routh-Hurwitz conditions for stability (see Paul Samuelson, p. 432) are

$$a_0>0$$
$$a_1>0$$
$$a_2>0$$
$$a_3>0$$
$$a_1a_2-a_0a_3>0$$

The condition $a_2>0$ is implied by the other conditions. Clearly the restrictions on the partial derivatives of E imply $a_3>0$. Furthermore if $E_y<1$ then $a_1>0$. The remaining condition, $a_1a_2-a_0a_3>0$, reduces to

$$E_pP^*+E_xA_x<-\frac{E_pP^*A_x}{A_y(E_y-1)}<0$$

This condition is more restrictive than the condition Tobin labels "[t]the critical necessary condition" (p. 199). Tobin's condition is

$$E_pP^*+E_xA_x<0$$

which, while sufficient to ensure that $a_2>0$, is *not* sufficient to ensure that $a_1a_2-a_0a_3>0$, and thus *not* sufficient for stability.

Inspection of the true condition reveals that the greater is A_y, the speed of adjustment of output, and the smaller is E_y, the marginal propensity to consume, the more likely is the W-K-P model stable. However, the effect of E_p on stability is ambiguous. Only if

$$\frac{A_x}{A_y(1-E_y)}<1$$

is Tobin's conclusion that a "strong negative price-level effect on aggregate demand... [is] conducive to stability" (pp. 199–200) correct. On the other hand Tobin's conclusions that "a weak price-expectation effect, and a slow response of price expectations to experience are conducive to stability" (pp. 199–200), are correct.

II

Tobin's M model is

$$\Pi=B_p(E-Y)+x$$
$$DY=B_y(Y^*-Y)$$
$$Dx=A_x(\Pi-x)$$

The inflation equation captures the idea that price, rather than quantity, reacts to differences between aggregate demand and aggregate output. As noted above, the output adjustment equation is separable from the rest and implies that output always adjusts to the equilibrium level, *regardless of the magnitude of the other variables of the model.* In particular output is independent of price and expected price. Now this is not inconsistent with Tobin's description of the Marshallian adjustment process (as quoted above) *if it is assumed that the supply curve for output of each firm is vertical.* However if the expected rate of inflation adjusts towards the actual rate by the Cagan learning hypothesis, then it is reasonable to argue, as it has been argued at length in the literature (see Edmund Phelps et al.), that the firm's *ex ante* supply curve might have a less than vertical slope. That is to say, if expectations are slow to adjust, a firm can increase employment and output at least temporarily in response to an increase in demand. This suggests respecifying the output adjustment equation as

$$Y-Y^{*}=B_y(\Pi-x)$$

If the actual rate of inflation exceeds (falls short of) the expected rate, aggregate output will be greater (less) than the "full-employment" level of output. This equation incorporates in dynamic terms the dependence of output adjustment on price which is emphasized in the Marshallian adjustment process. The expected rate of inflation affects wages and thus supply price whilst the actual rate affects the demand price.

The proposed M system is

$$\Pi=B_p\{E(P,x,Y)-Y\}+x$$

$$Y-Y^{*}=B_y(\Pi-x)$$

$$Dx=A_x(\Pi-x)$$

The steady-state solution is where

$$Dx=x=\Pi=0$$

This implies the natural rate of unemployment equilibrium solution

$$E=Y=Y^{*}$$

To analyze whether market forces will lead the system to the equilibrium solution, we can re-express the system as the following two-differential-equation system

$$D\Pi=\frac{DY}{B_y}+A_x\frac{(Y-Y^{*})}{B_y}$$

$$0=B_p\left\{E\left(P,\Pi-\frac{Y-Y^{*}}{B_y},Y\right)-Y\right\}-\frac{Y-Y^{*}}{B_y}$$

The characteristic equation is

$$0=D^2\left[B_p(E_y-1)-\frac{1}{B_y}\right]+D\left[\frac{B_p}{B_y}(P^{*}E_p+A_xE_x)\right]+\frac{B_pP^{*}A_xE_p}{B_y}$$

The necessary and sufficient condition for stability is

$$P^{*}E_p+A_xE_x<0$$

and $$E_y<1+\frac{1}{B_yB_p}$$

The first part of this condition is identical to the condition Tobin derives for his M model. However, there is an important difference in interpretation. Tobin's condition only applies to the price level; output in Tobin's model is stable, that is, approaches the equilibrium level, whatever the size of the adjustment parameters and the partial derivatives. In the M model put forward here, the stability condition refers to output as well as the price level. Thus it is possible for the M model to generate unemployment if the stability condition is not met.

III

A comparison of the two systems reveals several interesting points.

1) The level of output in either the W-K-P or the M model can be unstable. Thus the difference between the stability implications of the two types of adjustment processes is not as clear cut as suggested by Tobin's treatment. However, Tobin's general conclusion that "with stable monetary and fiscal policy, combined with price and wage flexibility, the adjustment mechanism of the economy may be too weak to eliminate persistent unemployment" (pp. 201–02), is strengthened since it follows not only if entrepreneurs follow the Walrasian adjustment process, but also if they follow the Marshallian adjustment process. Thus the analysis of this paper strengthens Tobin's case against fixed rules of monetary growth and fiscal policy.

2) The role of the marginal prospensity to consume in the stability conditions deserves comment. For the Marshallian formulation stability is possible even if the marginal prospensity to consume is greater than one—indeed the slower is the speed of adjustment of output (B_y) and of price (B_p), the greater the margin over unity which the marginal prospensity to consume can take without causing instability. For the W-K-P model, the maximum possible value of the marginal prospensity to consume consistent with stability is less than one. It is also interesting to observe that the quantity adjustment coefficient (A_y) enters the stability condition of the Walrasian model but that the *smaller* is this coefficient the more unstable the economy.

3) The possibility of instability in either model can be related to the expectations formation mechanism. If expectations are assumed to be determined exogenously, then $A_x=0$ and both models would be stable whatever the size of the other functional relations. It is also the case that the "Phillips loops" generated by an exogenous cycle in output can be related to this expectational parameter. In the W-K-P model

$$D\Pi = A_x A_p (Y-Y^*) + A_p DY$$

In the M model

$$D\Pi = \frac{A_x}{B_y}(Y-Y^*) + \frac{DY}{B_Y}$$

At the cyclical peak in output, $DY=0$ and $Y-Y^*>0$, implying $D\Pi>0$ in both models and thus clockwise Phillips loops: for the W-K-P model, the same point has been made by Frank Brechling and by Herschel Grossman. I have shown elsewhere that, for several versions of the Walrasian model, those versions which yield clockwise loops can be unstable while those which yield counterclockwise Phillips loops are stable. The models studied here give further support to the proposition that there is a link between the nature of the Phillips loops and the possibility of stability.

4) Alfred Marshall modified his dynamic structure by introducing the concept of the normal price (see pp. 374–77). The effect of excess demand on changes in the price level is mitigated to some extent by the difference between the actual price and the entrepreneur's concept of the normal price. The latter is not the same as the expected price—it is a longer-run concept and may be regarded as the average price over the trade cycle (see Edward Kuska for an application of this concept to the labor market). Since the money supply and the fiscal balance are held constant over the cycle in the models studied in this paper, the concept of the normal price can be included by respecifying the price adjustment relation as

$$\Pi = B_p(E-Y) + x - k(P-P_0)$$

where P_0 = "normal" price. Replacing the price adjustment equation in the proposed M system with the above the crucial stability condition becomes

$$P^*E_p + A_x E_x - \frac{P^*k}{B_p} < 0$$

Clearly this is a less stringent condition and so the introduction of the damping factor of normal price is a stabilizing force.

REFERENCES

F. Brechling, "The Trade-off between Inflation and Unemployment," *J. Polit. Econ.*, July/Aug. 1968, *76*, Part II, 712–37.

M. Friedman, "A Theoretical Framework for Monetary Analysis," *J. Polit. Econ.*, Mar./Apr. 1970, *78*, 193–238.

H. I. Grossman, "The Cyclical Pattern of Unemployment and Wage Inflation," *Economica*, Nov. 1974, *41*, 403–13.

E. A. Kuska, "The Simple Analytics of the Phillips Curve," *Economica*, Nov. 1966, *33*, 462–67.

Axel Leijonhufvud, *On Keynesian Economics and the Economics of Keynes*, London 1968.

I. McDonald, "The Dynamic Interactions between the Phillips Relation and the IS-LM Model," *Econ. Rec.*, forthcoming.

Alfred Marshall, *Principles of Economics*, 8th ed., London 1920.

Edmund S. Phelps et al., *The Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

Paul A. Samuelson, *Foundations of Economic Analysis*, Cambridge, Mass. 1947.

J. Tobin, "Keynesian Models of Recession and Depression," *Amer. Econ. Rev. Proc.*, May 1975, *65*, 195–202.

NOTES

The Professional Placement Service will be held December 28–30, 1980, at the Dallas Hilton, Dallas, Texas. Operating hours will be December 28, 10:00 A.M.–5:00 P.M.; December 29, 9:00 A.M.–5:00 P.M.; December 30, 9:00 A.M.–2:00 P.M. Special room rates have been established: $26/single; $30/doubles and twins; $50/one-bedroom suites; $75/two-bedroom suites. Hotel reservation cards will be mailed upon receipt of request for placement forms.

Request for placement forms should be made to Ms. Janet Aschenbrenner, National Registry for Economists, Illinois Job Service, 40 West Adams Street, Chicago, IL 60603 no later than November 15. Completed forms *must* be returned by December 1. You do not have to attend to have your listing on file. There is no registration fee.

Since 1946, Joseph A. Livingston, now associated with the *Philadelphia Inquirer*, has been conducting a systematic survey of economic forecasts by a small group of business and academic economists. The cross-section responses to these survey questions over the period 1946–79 are now available from the Research Department of the Federal Reserve Bank of Philadelphia. Researchers interested in obtaining the Livingston data set on magnetic tape or desiring additional information should contact Carole Moeller, Research Department, Federal Reserve Bank of Philadelphia, 100 North Sixth Street, Philadelphia, PA 19106. There will be no charge provided the tape is returned to the Federal Reserve Bank of Philadelphia.

Call for Papers; The annual meeting of the Southwestern Finance Association will be held March 5–8, 1981, at the Fairmont Hotel, New Orleans, Louisiana. Persons wishing to present a paper should submit a two-page abstract by September 15. Those wishing to participate as a discussant or session chairman should indicate their areas of interest in a brief letter. All topics of finance will be considered, including real estate, special decision science approaches to financially oriented problems, and other issues related to accounting and finance. Send material to Professor Stanley Block, Program Chairman, M. J. Neeley School of Business, Texas Christian University, Fort Worth, Texas 76129.

Call for Papers: The annual meeting of the History of Economics Society will be held June 1–3, 1981, at Kellogg Center, Michigan State University. Suggestions for topics and speakers are invited, as are proposals for papers. Please send to the President-Elect Warren J. Samuels, Department of Economics, Michigan State University, East Lansing, MI 48824. Please submit before December 1, 1980.

The Secretary of the American Economic Association calls attention to the policy adopted by the Executive Committee in 1980 of granting complimentary memberships to all members sixty-eight years of age or older who have retired from their professional activities, provided they have been members of the Association for twenty years or more. These complimentary members do not receive the Association's publications but are entitled to all other rights and privileges of membership.

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United State, Mexico, and Canada that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to air fare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for applications to be received in the ACLS office are: meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 345 East 46th St, New York, NY 10017, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting. Even when plans are incomplete, a prospective applicant should request forms in advance of the cut-off date, since deadlines are firm and no exceptions are permitted. Awards will be announced approximately two months after each deadline.

The Eleventh International Atlantic Economic Conference will be held at the London School of Economics, August 14–21, 1981. The theme will be "International Kaleidoscope." Those wishing to present a paper must submit two copies of a 500-word summary prior to November 1, 1980, with a separate cover sheet giving

(1) Name/Location of conference; (2) Name(s) of authors; (3) Institution or affiliation; (4) Full mailing address; (5) Telephone number; (6) Number and name of primary category of the paper. A submission fee of $30.00 (U.S.) is required. Those wishing to be discussants or chairmen should also submit to John M. Virgo, International Program Chairman, Atlantic Economic Conference, Box 258, Worden, IL 62097.

The National Humanities Center will admit approximately forty Fellows for the academic year 1981–82. Fellowships are available for Young Fellows (scholars three to ten years beyond the doctorate), Senior Fellows (scholars more than ten years beyond the doctorate), and Fellows for the following special seminars: "Energy and the Values of Modern Society"; "The Idea of a Profession"; "The Charles Frankel Seminar on Citizenship." The Center will provide full support to some Fellows at the level of their normal academic salaries and encourages application for residence from Fellows who have full or partial support from other sources. The deadline for applications is January 10, 1981, except for scholars from European nations, who should apply by December 10, 1980, to Dr. Raymond Georis, European Cultural Foundation, 51 Rue de la Concorde, 1050 Brussels, Belgium. All others may obtain information and applications from the National Humanities Center, P.O. Box 12256, Research Triangle Park, North Carolina 27709.

The Division of Science Resources Studies of the National Science Foundation announces the continuation of its Program for the Analysis of Science Resources. Grants will be awarded for studies focusing on aspects of training and employment of scientific and technical personnel, funding of scientific and technical activities, and outputs and impacts of scientific and technical activities. Proposals may be submitted at any time; however, they should be received by November 14, 1980 to insure consideration for funding in the 1981 fiscal year. Most awards are expected to be for one to two person-years of effort. You may obtain the announcement for the *Program for the Analysis of Science Resources* (NSF 80-19) from Division of Science Resources Studies, National Science Foundation, 1800 G Street NW, Washington, D.C. 20550.

Harvard Law School offers four or five Liberal Arts Fellowships to college and university teachers in the arts and sciences for the academic year, 1981–82. The purpose of the fellowships is to enable teachers in the social sciences or humanities to study fundamental techniques, concepts, and aims of law. Fellowship holders will take at least two first-year courses in law, in addition to more advance courses, and will participate in a joint seminar. The year of study will not count toward a degree. The fellowship grant covers tuition and health fee only. Applications should include a biographical resume (including academic record and list of publications), a statement explaining what the applicant hopes to achieve through his year of study, and two letters of recommendation (mailed to the Chairman directly from the referees). There is no special application form. The deadline for 1981–82 is January 15, 1981. Submit applications to the Chairman, Committee on Liberal Arts Fellowships, Harvard Law School, Cambridge, MA 02138. Awards will be announced before February 15, 1981.

On December 3–7, 1980, the Institute for Democratic Socialism will hold a conference, "Eurosocialism and America: An International Exchange," at the Capital Hilton, Washington, D.C., funded by the German Marshall Fund of the United States. For more information, contact Nancy Lieber, Conference Director, Institute for Democratic Socialism, 853 Broadway, Suite 801, New York, NY 10003 (telephone 212+260-3270).

The 1979 winner of the Leonard J. Savage Thesis Award for an outstanding thesis in Bayesian econometrics and statistics was Kevin James McConway for his dissertation, "The Combination of Experts' Opinions in Probability Assessment: Some Theoretical Considerations," completed at the University College London. The Thesis Evaluation Committee included Edward E. Leamer, Chairman, Christopher B. Barry, Richard E. Kihlstrom, John W. Pratt, and Arnold Zellner (*ex officio*).

Deaths

Lubomir A. D. Dellin, professor of economics, University of Vermont, Mar. 9, 1980.

Bob R. Holdren, professor of economics, Iowa State University, Jan. 1, 1980.

Chester Morgan, professor of economics, University of Iowa, Feb. 29, 1980.

Michael L. Visscher, assistant professor of economics, Ohio State University, Feb. 20, 1980.

Retirements

Paul G. Craig, professor of economics, Ohio State University, June 1980.

David B. McCalmont, associate professor of economics, Ohio State University, June 1980.

Herbert S. Parnes, professor of economics, Ohio State University, Aug. 1980.

David I. Siskind, assistant division chief, Construction Statistics Division, Bureau of the Census, Feb. 29, 1980.

Promotions

Esmond T. Adams; professor, department of economics, Central Missouri State University, Sept. 1, 1980.

Patrick J. Corcoran: chief, Business Conditions Division, domestic research department, Federal Reserve Bank of New York, Mar. 13, 1980.

Paul H. Engelmann: associate professor, department of economics, Central Missouri State University, Sept. 1, 1980.

William C. Melton: chief, Monetary Analysis Division, domestic research department, Federal Reserve Bank of New York, Feb. 28, 1980.

Barron A. Peake: associate professor of economics, Golden Gate University, July 1979.

Rao N. Singamsetti: associate professor of economics, Austin Dunham Barney School of Business, University of Hartford, Sept. 1, 1980.

John A. Wenninger: research officer and senior economist, Research and Statistics Function, Federal Reserve Bank of New York, Mar. 1, 1980.

Administrative Appointments

William A. Fitzsimmons: assistant vice president, systems planning, New York Stock Exchange, Product Development Division, Apr. 11, 1980.

William E. Gibson: senior vice president, McGraw-Hill, Inc., Mar. 31, 1980.

Katherine B. Magrath: director of equity investments, Ford Foundation, March 1980.

Barron A. Peake: director, Executive MBA Program, Golden Gate University, July 1979.

Mark B. Schupack: associate provost, Brown University, July 1, 1980.

Ronald Tiersky, Amherst University: director, Bologna Center, School of Advanced International Studies, Johns Hopkins University, 1980–81.

Bernard L. Weinstein: associate director for federal affairs, Southern Growth Policies Board, Washington, D.C., Feb. 1980.

Appointments

Richard G. Anderson, Michigan State University: assistant professor, department of economics, Ohio State University.

Ernst R. Berndt, University of British Columbia: professor of applied economics, The Sloan School, M.I.T., July 1, 1980.

John Boschen, Brown University: economist, Board of Governors of the Federal Reserve System, Sept. 1980.

Stanford Calderwood: visiting professor, economics department, Wellesley College, Jan. 1981.

Donald C. Cox: economist, Business Conditions Division, domestic research department, Federal Reserve Bank of New York, Jan. 16, 1980.

John B. Chilton, Brown University: lecturer, department of economics, Faculty of Social Sciences, University of Western Ontario, Sept. 1980.

Evangelos M. Falaris, University of North Carolina-Chapel Hill: assistant professor, department of economics, Ohio State University.

Robert Gough: visiting professor, economics department, Wellesley College, Sept. 1980.

Marvin L. Hayenga: associate professor, department of economics, Iowa State University, Jan. 22, 1980.

Richard W. Heaps: assistant professor of economics, University of Vermont, Sept. 1980.

Michael Horrigan: assistant professor of economics, Williams College, July 1, 1980.

Richard A. Jensen, Northwestern University: assistant professor, department of economics, Ohio State University.

Clive V. Jones: instructor in economics, Division of Social Sciences, University of Colorado-Denver, Jan. 1979.

Edward C. Kienzle, Central Connecticut State College: assistant professor of economics, Stonehill College, Sept. 1, 1980.

Nancy J. Kimelman, Occidental College: economist, research department, Federal Reserve Bank of Dallas, Sept. 1980.

David Lindauer: assistant professor, economics department, Wellesley College, Jan. 1981.

John J. Merrick, Jr., Brown University: assistant professor of finance, Graduate School of Business Administration, New York University, Sept. 1980.

Len Nichols: instructor, economics department, Wellesley College, Sept. 1980.

Diann Painter: associate professor, economics department, Wellesley College, June 1980.

Michael T. Rabbitt, Washington University (St. Louis): assistant professor, department of economics, Central Missouri State University, Sept. 1, 1980.

Susan Randolph: instructor, economics department, Wellesley College, Jan. 1981.

Jess A. Robinson: instructor, department of economics, Iowa State University, Feb. 15, 1980.

Elyce Rotella: assistant professor, economics department, Wellesley College, Sept. 1980.

Morton O. Schapiro: assistant professor of economics, Williams College, July 1, 1980.

Joyce Shackett: assistant professor, department of economics, University of Missouri-Columbia, Sept. 1, 1980.

Steven Swidler, Rice University: assistant professor of business administration, University of Wisconsin-Milwaukee, Sept. 1980.

Rona Wilensky: assistant professor of economics, Williams College, July 1, 1980.

Richard J. Wilike, Johns Hopkins University: assistant professor, department of economics, Ohio State University.

Arthur G. Woolf: assistant professor of economics, University of Vermont, Sept. 1980.

Resignations

Charles C. Cox, Ohio State University: Texas A&M University, Sept. 1980.

Robert A. Driskill, Ohio State University: University of California-Davis, June 1980.

Tatsuro Ichiishi, Carnegie-Mellon University, Aug. 25, 1980.

Editor's Note: The editorial office of the *American Economic Review* will move to the University of California–Los Angeles, January 1, 1981. The new managing editor will be Professor Robert W. Clower. Starting October 1, 1980, please address all new manuscripts to the *American Economic Review*, Editorial Office, University of California, Los Angeles, CA 90024.

Correspondence with regard to manuscripts submitted before October 1 should be sent to the present editorial office at Brown University, Providence, RI 02912.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

CALL FOR PAPERS

1981 MEETINGS

of the

AMERICAN ECONOMIC ASSOCIATION

Members wishing to give papers or to make suggestions for the program of the meetings to be held in Washington, D.C., December 28-30, 1981, are invited to write to Professor Gardner Ackley, Department of Economics, The University of Michigan, Ann Arbor, MI 48109. As in the past, the sessions sponsored by the American Economic Association will be of two types: invited papers and contributed papers. Most of the invited papers will be published in the *Papers and Proceedings* issue of the *American Economic Review* of May 1982; the contributed papers will not be so published. Suggestions of topics or proposals of papers for the invited sessions should be submitted as soon as possible. Abstracts of proposed contributed papers must be received no later than March 1, 1981. Econometric studies or highly mathematical papers are not appropriate for the sessions sponsored by the American Economic Association. However, members wishing to present econometric papers may submit their proposals or abstracts to the Econometric Society which meets with the American Economic Association and normally schedules a number of sessions involving such papers.

Of Economic Importance...

New 1981 Titles From P-H

THE ECONOMIC PROBLEM, Sixth Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 672 pp. (est.) Cloth

UNDERSTANDING MACROECONOMICS, Seventh Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 384 pp. (est.) Paper

UNDERSTANDING MICROECONOMICS, Fifth Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 352 pp. (est.) Paper

MONEY, BANKING AND ECONOMIC ANALYSIS, Second Edition
Thomas D. Simpson, George Washington University
1981 502 pp. (est.) Cloth

ECONOMICS OF THE FIRM: Theory and Practice, Third Edition
Arthur A. Thompson, Jr., Univ. of Alabama
1981 656 pp. (est.) Cloth

MARKETS, DECISIONS, AND ORGANIZATIONS: Intermediate Microeconomic Theory
Roger A. McCain, Temple University
1981 620 pp. (est.) Cloth

BUSINESS, GOVERNMENT, AND THE PUBLIC, Second Edition
Murray L. Weidenbaum, Washington University
1981 336 pp. (est.) Cloth

INTERMEDIATE MACROECONOMICS
William A. Kelly, Jr., University of North Carolina – Chapel Hill
1981 500 pp. (est.) Cloth

BENEFIT-COST ANALYSIS OF GOVERNMENT PROGRAMS
Edward M. Gramlich, Univ. of Michigan
1981 352 pp. (est.) Cloth

MACROECONOMICS: Theory, Evidence and Policy, Second Edition
Frank C. Wycoff, Pomona College
1981 550 pp. (est.) Cloth

COPING WITH CAPITALISM: The Economic Transformation of the United States, 1776-1980
Roger L. Ransom, University of California – Riverside
1981 256 pp. (est.) Paper

MACROECONOMICS FOR MANAGEMENT
Mary Louise Hatten, Boston College
1981 384 pp. (est.) Cloth

REFORMING REGULATION: Processes and Problems
Lawrence J. White, New York University
1981 128 pp. (est.) Cloth

For further information, or to order or reserve examination copies, please write: Robert Jordan, Dept. J-668, Prentice-Hall, Inc., Englewood Cliffs, NJ 07632
For "SUPER-QUICK" Service, dial TOLL FREE (800) 526-0485,* between 9:00 a.m.-4:30 p.m. EST, Mon.-Fri.
* not applicable in New Jersey, Alaska, Hawaii, or Puerto Rico

Prentice-Hall

New from Princeton

The Economic Crisis and American Society

MANUEL CASTELLS

A critical analysis of America's society and economy which identifies the new forms of the capitalist mode of production and the contradictory nature of its class relations as the root of the country's problems. Taking a hard look at the recent crisis afflicting Western economies, Manuel Castells suggests that the very structures that fostered economic growth since 1945 are the same structures that are now undermining these economies.
Cloth, $20.00. Paperback, $5.95.

Female Labor Supply

Theory and Estimation

Edited by JAMES P. SMITH

This collection of original essays brings econometric theory to bear on the problem of estimating the labor force participation of women. Five scholars here examine, both theoretically and empirically, the determinants of women's wages in the market, the value of their home time, and the factors that affect their employment. *A Rand Corporation Research Study.* $20.00

Human Resources in Japanese Industrial Development

SOLOMON B. LEVINE and HISASHI KAWADA

By focusing on the educational and skill training institutions Japan had developed to generate human resources for modern industry, this book represents a new contribution to the historical analysis of Japan's modern economic growth. $20.00

Reasoning About Discrimination

The Analysis of Professional and Executive Work in Federal Antibias Programs

RICHARD A. LESTER

Developing an analytical basis for judging sex, race, and age discrimination in professional and executive employment, Lester examines federal antibias programs applied in the 1970s, criticizes the analytical underpinning of the programs, and suggests means for improving their suitability for professional and executive workers. $17.40

Multinationalism, Japanese Style

The Political Economy of Outward Dependency

TERUTOMO OZAWA

"This is a valuable analysis and brief history of Japanese overseas direct investment in the postwar period. The links between private and public activity, trade policy and the changing structure of Japanese industry, export promotion and the effort to obtain secure supplies of raw materials and oil are well brought out."
—*Foreign Affairs.* $16.50

Princeton University Press
Princeton, New Jersey 08540

M. E. Sharpe, Inc.

Publisher of CHALLENGE and
Journal of Post Keynesian Economics

NEW CLASSROOM PAPERBACKS

Economic Democracy
The Challenge of the 1980s

By Martin Carnoy and
Derek Shearer

"This book leaps beyond the 'conventional wisdom' of Establishment economists and goes to the heart of America's problems..."
—*Bennett Harrison,*
Massachusetts Institute of Technology

Cloth $15.00 Paper $7.95*

A Guide to Post-Keynesian Economics

Edited by Alfred S. Eichner

"... the essential elements of a major development in contemporary economics that has not received the attention it deserves from mainstream economists." —*Choice*

Cloth $12.95 Paper $4.95*

The Megacorp and Oligopoly
Micro Foundations of Macro Dynamics

By Alfred S. Eichner

"...Without doubt this is a stimulating and well written book which deserves a wide audience."
—*Economic Journal*

Paper $8.95

*Distributed by Pantheon Books

Unemployment & Inflation
Institutionalist and Structuralist Views

Edited by Michael J. Piore

"(Piore's) five essays and those of the other contributors provide an excellent introduction to the institutionalist views of the economy."
—*Richard C. Edwards,*
University of Massachusetts

Cloth $20.00 Paper $7.95

The Yugoslav Economic System
The First Labor-Managed Economy in the Making

By Branko Horvat

"This work, by Yugoslavia's best-known economist, is now the best and most concise work available describing that country's unique market socialist economy..."
—*Sociology Reviews of New Books*

Paper $7.95

The Advanced Capitalist System
A Revisionist View

By Lynn Turgeon

"... (The author's) insights into economic dynamics are very unorthodox and stimulate one to think. ... A very exciting and useful book." —*Howard J. Sherman,*
University of California, Riverside

Cloth $15.00 Paper $7.95

M. E. Sharpe, Inc. 901 N. BROADWAY, WHITE PLAINS, N.Y. 10603

Please mention The American Economic Review *When Writing to Advertisers*

For the best in principles texts—

The first brand-new text for the 80's—

ROGER N. WAUD'S

ECONOMICS

Recent shocks to the supply side of the economy, such as the rapid rise in imported oil prices, are making it painfully apparent that aggregate supply must be worked into macroeconomic analysis.

"This text shows the relationship between total demand and supply and illustrates why it is possible to have increases in the general price level and the unemployment rate at the same time. After standard Keynesian income analysis has been developed fully, the supply side of the economy is carefully integrated with it. This not only facilitates a comparison of Keynesian and Monetarist views, but also helps instructors deal candidly with cost-push inflation and the effects of stagflation on monetary and fiscal policy.

"Students expect, and should receive, a realistic analysis of such problems in an introductory course. We owe it to them."—Roger N. Waud

January 1980. 848 pages. ISBN: 0-06-046963-3. Instructor's Manual with Test Bank, Study Guide, Transparency Masters, and a computer-based examination system—ACCESS.

Two-volume paperback edition also available: *Macroeconomics,* 512 pages; and *Microeconomics,* 448 pages.

Start off 1981 with the NEW sixth edition of—

***Lipsey & Steiner's* ECONOMICS**

To request examination copies, write to Suite 5D, Harper & Row, 10 East 53d St., New York, N.Y. 10022. Please include course title, enrollment, and present text.

Harper & Row—the leader in the field.

To meet the challenges of the 80's—

THE ECONOMICS OF PUBLIC ISSUES

Fifth Edition

by DOUGLASS C. NORTH & ROGER LEROY MILLER

Ideal as a supplement to principles of economics courses, this fully updated edition offers 34 concise and stimulating original essays. The fifth edition includes 18 new chapters on everything from drug regulation, revival of the draft, and illegal aliens to wage and price controls, black economic progress, and urban sprawl. The popular chapters on agriculture, medical care, safer products, and others are thoroughly updated.

Carefully planned to parallel the development of basic microeconomic theory in leading texts, chapters are now grouped under major subject categories.

January 1980. 214 pages. ISBN 0-06-044875-X. Paper. Instructor's Manual.

The latest edition of an economics classic—

ECONOMICS TODAY

Third Edition

ROGER LEROY MILLER

A clearly written, practical revision of a complete teaching/learning package for the introductory one-, two-, or three-term principles course. Provides a solid understanding of basic economic tools and theory and presents many practical applications through topical issues.

1979. 747 pages. ISBN 0-06-385481-3.

Instructor's Manual. Student Learning Guide. Test Item File. Transparency Masters. Independent Study Audio Cassette. Modules and Study Guides.

Harper & Row

EXECUTIVE SEARCH

SPECIALISTS IN:

ECONOMICS
ECONOMETRICS
CORPORATE PLANNING
STRATEGIC PLANNING
MARKETING
OPERATIONS RESEARCH

STONEHILL IS AN EXECUTIVE SEARCH FIRM STAFFED WITH A TEAM OF PROFESSIONALS, EXPERIENCED IN SEARCH AND RECRUITMENT TECHNIQUES.

WE TAKE PRIDE IN OUR NATIONWIDE REPUTATION FOR HIGH STANDARDS OF PROFESSIONALISM IN EXECUTIVE RECRUITING.

OUR CLIENTS ARE "FORTUNE 500" COMPANIES, MAJOR CORPORATIONS, FINANCIAL INSTITUTIONS, AND CONSULTING FIRMS.

CONTACT: MR. GARY J. STEVENS, PRESIDENT
MS. DEE DROST, V.P. EXECUTIVE SEARCH

STONEHILL
MANAGEMENT CONSULTANTS, INC.
521 FIFTH AVENUE
NEW YORK, NEW YORK 10017

212 490-1776 **SUITE 1160**

PRAEGER

entering our fourth decade of distinguished publishing

MANAGING GROWTH IN THE 1980's
Toward a New Economics
ROBERT D. HAMRIN, Environmental Protection Agency
318 pp. 1980 ISBN 0-03-054061-5 $23.95
Paperback available! ISBN 0-03-054056-9 $9.95

GLOBAL BUSINESS STRATEGY FOR THE 1980s
F.T. HANER, University of Delaware
244 pp. 1980 ISBN 0-03-047196-6 ca. $24.95

THE ETHICAL BASIS OF ECONOMIC FREEDOM
edited by **IVAN HILL,** American Viewpoint, Inc.
Foreword by **VERMONT ROYSTER,** *The Wall Street Journal*
"It is an impressive volume and merits the attention it is getting." — U.S. News & World Report
"It transcends any investment advice." — Financial World
ca. 272 pp. September 1980 ca. $9.95

PRAEGER
Praeger Publishers
CBS Educational and Professional Publishing
521 Fifth Avenue, New York, New York 10017
(212) 599-8400

STAFF PAPERS

Published by the International Monetary Fund in March, June, September, and December. Summaries of each paper in French and Spanish, as well as in English.

Vol. 27 June 1980 No. 2

Fiscal Content of Financial Programs Supported by Stand-By Arrangements in the Upper Credit Tranches, 1969–78 ... W.A. BEVERIDGE and MARGARET R. KELLY

Monetary Shocks and the Dynamics of Inflation .... MOHSIN S. KHAN

The World Model of Merchandise Trade: Simulation Applications
DUNCAN M. RIPLEY

Short-Run Effects of Exchange Rate Changes on Terms of Trade and Trade Balance
ERICH SPITÄLLER

Gasoline Taxation in Selected OECD Countries, 1970–79
ALAN A. TAIT and DAVID R. MORGAN

The International Monetary Fund, 1978–79: A Selected Bibliography
ANNE C. M. SALDA

Subscription: US$9.00 a volume; US$3.00 a single issue. Special rate to university libraries, faculty, and students: US$4.00 a volume; US$1.00 a single issue.

Advice on payment in other currencies will be given upon receipt of order.

The Secretary, Attention: Publications Section, International Monetary Fund, Washington, D.C. 20431 U.S.A.

MIT
ECONOMICS
New books

THE ORIGINS AND DEVELOPMENT OF LABOR ECONOMICS
by Paul J. McNulty
$17.50

THE DEFENSE INDUSTRY
by Jacques S. Gansler
$19.95

GETTING STARTED:
The Youth Labor Market
by Paul Osterman
$20.00

BLACK WOMEN IN THE LABOR FORCE
by Phyllis A. Wallace
with the assistance of
Linda Datcher and
Julianne Malveaux
$15.00

THE COLLECTED PAPERS OF FRANCO MODIGLIANI
Volume 1:
Essays in Macroeconomics
Volume 2:
The Life Cycle Hypothesis of Saving
Volume 3:
The Theory of Finance and Other Essays
edited by Andrew Abel
$25.00 each volume

COLLECTED ECONOMIC PAPERS OF JOAN ROBINSON
Five Volumes
$25.00 each volume
Boxed set, with separately bound index, $125.00

Paperbacks

NORTH-SOUTH
A Program for Survival
A Report of the Independent
Commission on International Development Issues
under the Chairmanship of
Willy Brandt
$4.95

THE ECONOMIC THEORY OF POLLUTION CONTROL
by Paul Burrows
$8.95; cloth $25.00

A GUIDE TO ECONOMETRICS
by Peter Kennedy
$7.95; cloth $19.95

THE INNOVATION DECISION IN SOVIET INDUSTRY
by Joseph S. Berliner
$15.00

The MIT Press
Massachusetts Institute of
Technology
Cambridge, Massachusetts 02142

HARVEY J. LEVIN

Fact and Fancy in Television Regulation

An Economic Study of Policy Alternatives

HOW DIVERSE CAN, AND SHOULD, TV programming be? And especially, in what precise ways does governmental regulation of TV affect (or fail to affect) the programs station owners produce —programs which, in the final analysis, shape in such large measure the values of Americans? It is to these timely and beguiling questions that Harvey Levin addresses his dispassionate assessment of the complex relationship between government and the TV industry. Analyzing data drawn from the history of the FCC's regulatory decisions, as well as from interviews with numerous government and industry officials, Professor Levin shows how the present form of restrictive governmental regulation almost always results in higher profits and rents for TV stations, with no concomitant increase in programming diversity.

In addition, Professor Levin investigates various other aspects of the media market, from the particular kinds of crucial decisions that are made when, for example, a newspaper owns a TV station, to the kinds of problems that arise when commercial rents are taxed to fund public TV; from the brand of programming we are offered when a monopoly controls a given TV market to the nature of programming in a situation of steady and fair competition. Following a comprehensive assessment, the author makes a compelling case for diversification of station ownership, in order to be "safe rather than sorry." He also argues for the entry of new stations, more extensive support of public TV, and some form of quantitative program requirements — all of which will help bring about greater program diversity.

Professor Levin's volume provides us with a fully documented and sharply focused analysis of the theories, policies, and problems of one of the most powerful and misunderstood of contemporary institutions.

Contents

1. Scope and Method

Part I: *Basic Framework*. 2. Economic Structure and Regulatory Framework. 3. The Concept and Measurement of Diversity. 4. Origin and Incidence of Economic Rents in Broadcasting. 5. Objectives versus Results in Federal Broadcast Regulation.

Part II: *A Critique of Current Policies*. 6. The Policy on Ownership Diversity. 7. The Cross-Ownership Proceeding: A Case Study in Factual Deficiencies. 8. Structural Diversity, Owner Diversity, and the Role of Public Television: A Comparative Assessment. 9. Alleged Bases of National Television Network Power. 10. The Limits of Network Regulation. 11. Objectives versus Consequences in Prime-Time Access.

Part III: *Assessment of Selected Alternatives*. 12. A Note on the Regulatory Implications of Scarcity Rents in Television. 13. Quantitative License Renewal Requirements. 14. The Use of Scarcity Rents to Fund Public Broadcasting.

Appendix I: Basic Methodological Issues. Appendix II: Principal Estimating Equations. Index.

About the author:

Harvey J. Levin is Professor of Economics at Hofstra University, and Senior Research Associate at the Center for Policy Research.

LC 79-90148 ISBN: 0-87154-531-4
544 pages **$20.00**

Distributed by
Basic Books, Inc.
10 East 53rd Street
New York, New York 10022

RUSSELL SAGE FOUNDATION

CAMBRIDG

Economics and Consumer Behavior

Angus Deaton and John Muellbauer

The authors have used the duality approach to provide an integrated and accessible treatment of the economics of consumer behavior. This book focuses on applications of the theory to welfare economics and econometric analysis. Much of the material is too recent to have appeared in any previous text. The structure of the book is designed to permit its use in a wide variety of courses at different levels, and examples are provided throughout.

Hardcover $49.50 Paper $18.95

Asset Markets, Exchange Rates, and Economic Integration

A Synthesis

Polly Reynolds Allen and Peter B. Kenen

Allen and Kenen have developed a new theoretical model for economies participating in international trade. They demonstrate that market integration fosters interdependence among nations in the conduct of economic policy. The authors apply the results of their work to a cost-benefit analysis of both monetary and fiscal unification, emphasizing the effect of the external exchange regime on the members of an economic union. $55.00

Inflation and the Personal Income Tax

An International Perspective

Vito Tanzi

A detailed analysis of the impact of inflation on income tax systems in many countries. Tanzi examines some of the more important distortions of income tax systems caused by inflation and discusses possible corrective measures, comparing them with the actual experiences of the United States, United Kingdom, France, Australia, Canada and other countries in their pursuit of policies aimed at neutralizing the impact of inflation. $27.50

Economics and Corporate Strategy

C. J. Sutton

Drawing on a wide range of economics and management literature, Sutton shows how an understanding of industrial economics can help in analyzing strategic decisions. He explains how a firm's development must be adapted to its environment, its history, and the experience of its personnel. Included are discussions of such topics as integration and diversity, the growing importance of multinational operations, the strategic role of mergers, and innovation.

Hardcover $27.50 Paper $7.95

The Collected Writings of John Maynard Keynes

Donald Moggridge, Editor

Volume 25: Activities 1940-44: Shaping the Post-War World: The Clearing Union *$35.00

Volume 26: Activities 1943-46: Shaping the Post-War World: Bretton Woods and Reparations *$35.00

Volume 27: Activities 1940-46: Shaping the Post-War World: Employment and Commodities *$35.00

** tentative*

Industrialization and the Basis for Trade

R. A. Batchelor, R. L. Major and A. D. Morgan

NIESR Economic and Social Studies 32

$39.50

The Economic Status of Australian Aborigines

Jon C. Altman and John P. Nieuwenhuysen

$29.50

Birds of Passage

Migrant Labor in Industrial Societies

Michael J. Piore *Now in Paperback* $6.95

The Power to Tax

Analytical Foundations of a Fiscal Constitution

Geoffrey Brennan and James Buchanan

A radically new approach to the understanding and evaluation of our fiscal system. The authors proceed from the question of how much "power to tax" a citizen would voluntarily grant to government under the terms of a social contract devising a fiscal constitution. They develop a model of government as a "revenue-maximizing Leviathan" and trace out the restrictions on the power to tax that might emerge from the citizen's constitutional deliberations. *$24.95

Theory of International Trade

A Dual, General Equilibrium Approach

Avinash Dixit and Victor D. Norman

Emphasizing that a trading equilibrium is general rather than partial and is often best modeled using dual or envelope functions, the authors give unified treatment of comparative statics and welfare and shed new light on the factor-price equalization issue. They also treat the modern specific-factor model in parallel with the usual Heckscher-Ohlin model. *Cambridge Economic Handbooks* *Hardcover $34.50 Paper $14.95

Capital Theory and Dynamics

Edwin Burmeister

Professor Burmeister focuses mainly on the pure role of time and the influence of future but certain events on the current economic state. He introduces such concepts as dynamic efficiency, Pareto optimality, the crucial role played by capital goods in intertemporal economies and portfolio problems. He discusses such issues as the "Cambridge controversies," Sraffa's Standard Commodity, competitive paths, saddlepoint instability, perfect foresight competitive equilibrium and rational expectations competitive equilibrium. *Cambridge Surveys in Economic Literature* *Hardcover $37.50 Paper $14.95

The Anti-Trust Laws of the U.S.A.

Third Edition

A. D. Neale and D. G. Goyder

Completely revised and updated, this third edition deals extensively with recent developments relating to mergers and acquisitions and to the question of how far U.S. jurisdiction should influence business practices outside the United States. It assesses the value and meaning of the antitrust system in relation to the body of anti-monopoly laws being developed by the EEC. *NIESR Publications 19* *Hardcover $39.50 Paper $15.95

Computers and Economic Planning
The Soviet Example
Martin Cave
Soviet and East European Studies $27.50

Socialist Economic Integration
Jozef M. van Brabant
Soviet and East European Studies $29.50

Dualism and Discontinuity in Industrial Societies
Suzanne Berger and Michael Piore $17.95

Profitability and Unemployment
Edmond Malinvaud
co-published with *Les Editions de la Maison des Sciences de l'Homme* $12.95

The Subsidized Muse
Public Support for the Arts in the United States
Dick Netzer
A Twentieth Century Fund Study
Now in Paperback $7.95

Cambridge University Press

32 East 57th Street, New York, N.Y. 10022

Books Published by
Oxford University Press for the World Bank

Policies for Industrial Progress in Developing Countries

Edited by JOHN CODY, UNIDO; HELEN HUGHES, The World Bank; and DAVID WALL, University of Sussex. Filling an important gap in the literature on industrial development, this collection of essays provides a handbook for administrators and a clearly written summary for students. The importance of the link between theory and administration is stressed and practical issues are given more weight than theoretical economic concepts.

1980 320 pp.; 4 tables cloth $16.95 paper $5.95

Inequality and Poverty in Malaysia
Measurement and Decomposition

SUDHIR ANAND, St. Catherine's College, Oxford. Anand's book is a sophisticated account of income inequalities and poverty in Malaysia which will be of particular interest to policy makers. The research is policy-oriented and the findings, to which his statistical technique is imaginatively applied, are thoroughly discussed.

Fall 1980 228 pp.; tables; formulae cloth $14.95 paper $5.95

Income Inequality and Poverty
Methods of Estimation and Policy Applications

NANAK KAKWANI, University of New South Wales. In this study, appropriate techniques have been developed to analyze the problems of size distribution of income using actual data, and these techniques have been applied to evaluate alternative fiscal policies affecting income distribution. The contributions include both systematization of existing knowledge as well as a number of new results.

1980 320 pp., graphs; charts cloth $19.95 paper $8.95

Migration in West Africa
The Demographic Aspects

K.C. ZACHARIAH, The World Bank; and JULIEN CONDÉ, Organization for Economic Cooperation. This work is a descriptive study of international, internal and rural-urban migration in nine West African countries. The volume and direction of external and internal migration are studied along with selected socio-economic characteristics of migrants and non-migrants. The study is based on the censuses of 1970 and 1976 and much of the demographic data used is entirely new.

October 1980 256 pp.; tables; maps cloth $16.95 paper $6.95

Models of Growth and Distribution for Brazil

LANCE TAYLOR, Massachusetts Institute of Technology; EDMAR BACHA and ELIANA CARDOSO, both of the University of Brazil; and FRANK LYSY, Johns Hopkins University. This work explores the Brazilian political economy. Using computable general equilibrium income distribution models, the authors focus on the interaction of growth and income distribution, both providing current ideas about the Brazilian economic model and pointing to future lines of research for politically observant economists.

1980 384 pp.; tables cloth $18.95 paper $7.95

Prices and publication dates are subject to change.

Oxford University Press
200 Madison Avenue, New York, New York 10016

A new edition of a leading text

International Economics

Second Edition

Bo Södersten, University of Lund

CONTENTS

Preface to the Second Edition

Introduction: International Economics and Economic Theory

Part I: The Pure Theory of International Trade • The Theory of Comparative Advantage and the Gains from Trade • Production Functions in International Trade: The Box Diagram • Comparative Advantage in the Heckscher-Ohlin Trade Model • Commodity and Factor Prices Under Trade: Factor-Price Equalization • The Gains from Trade and the Income Distribution • Factor Reversals and Factor Prices: Empirical Testing and the Leontief Paradox • Alternative Theories of Trade

Part II: Economic Growth and International Trade • Economic Growth and International Trade during the Nineteenth and Twentieth Centuries • A Model of Economic Growth and International Trade where Growth is Unspecified • Increases in Factor Endowments and International Trade: The Rybczynski Theorem • Technical Progress and International Trade • The Terms of Trade and the National Income in the Growth Process

Part III: Trade Policy • Tariffs under Optimal Market Conditions • Quotas and Quantitative Restrictions: State Trading • Tariffs, Subsidies, and Distortions in Commodity and Factor Markets • Import-Substitution versus Export-Push • GATT, the EEC, the Kennedy Round, and the Tokyo Declaration • UNCTAD and Trade Policies for Less Developed Countries • A New Economic World? • The Theory of Customs Unions • Direct Investments and the Multinational Firm

Part IV: The Balance of Payments and International Economic Policy • The Market for Foreign Exchange • The Balance of Payments • Foreign Trade and National Income • International Economic Policy and the Adjustment Mechanism • The Monetary Approach to the Balance of Payments • Fixed versus Flexible Exchange Rates • The Aims-Means Analysis for Reaching Internal and External Equilibrium

Part V: The International Monetary System • The Bretton Woods System • The Present International Monetary System • Notes and References

1980 524 pp., cloth $16.95

To request a complimentary examination copy please write (specifying course, present text and approximate enrollment) to:

St. Martin's Press P.O. Box 5352 New York, N.Y. 10163

St. Martin's Press

Recent Studies on

North-South Relations

published by or in cooperation with the Overseas Development Council

☐ **The Advanced Developing Countries: Emerging Actors in the World Economy**
by John A. Mathieson
Overseas Development Council, 1979
72 pp. **$3.00**

☐ **Positive Adjustment to a Changing World Economy**
by Florizelle B. Liser
Overseas Development Council, 1980
6 pp. **$ .25**

☐ **Co-Financing for Development: Why Not More?**
by Roger S. Leeds
Overseas Development Council, 1980
64 pp. **$3.00**

☐ **Policy Alternatives for a New International Economic Order: An Economic Analysis**
edited by William R. Cline
Praeger for ODC, 1979
410 pp. Paperback: **$7.95**

☐ **The United States and World Development: Agenda 1980**
by John W. Sewell and the Staff of the Overseas Development Council
Praeger for ODC, 1980
256 pp. Paperback: **$6.95**

☐ **An Alternative Road to the Post-Petroleum Era: South-North Cooperation**
by James W. Howe
and James Tarrant
Overseas Development Council, 1980
80 pp. **$5.00**

☐ **Mobilizing Technology for World Development**
edited by Jairam Ramesh
and Charles Weiss, Jr.
Praeger for ODC, 1979
240 pp. Paperback: **$6.95**

☐ **The Uncertain Promise: Value Conflicts in Technology Transfer**
by Denis Goulet
IDOC/North America for ODC, 1977
324 pp. Paperback: **$5.95**

☐ **Beyond the North-South Stalemate**
by Roger D. Hansen
McGraw-Hill. By special arrangement with the 1980s Project of the Council on Foreign Relations and McGraw-Hill, a limited paperback edition is available exclusively from ODC. 1979
348 pp. Paperback: **$5.95**

☐ **Complete Catalog of ODC Publications**

Please check the publications above you wish to receive and enclose this form with your check for the total amount **minus a 10% discount available with this coupon.**

$ enclosed.

Name______________________________

Address ______________________________

City ______________ **State**________ **Zip**______

overseas development council
1717 Massachusetts Ave., N.W., Suite 501B
Washington, D.C. 20036

ECONOMIC ANALYSIS OF INTERCITY FREIGHT TRANSPORTATION

George W. Wilson

This rigorous micro-economic analysis examines the criteria relevant to a more efficient use of transportation facilities in this country, present and future. Wilson places the transportation industries in the context of the economy as a whole.

384 pages, append., index, drawings $25.00

SOVIET ENERGY TECHNOLOGIES

Planning, Policy, Research and Development

Robert W. Campbell

Detailed discussions of all major Soviet energy capabilities. Focuses on the research and development structures and their effectiveness, the availability of resources, and comparisons with U.S. and other Western technologies. An excellent evaluation with broad implications for future development.

288 pages, index $22.50

EAST EUROPEAN INTEGRATION AND EAST-WEST TRADE

Edited by Paul Marer and John Michael Montias

Analytical and theoretical essays by leading authorities in the fields of international trade, economics, and East European affairs. Establishes a comprehensive, up-to-date, and balanced Western perspective on foreign trade in the centrally planned economies of Eastern Europe.

416 pages $22.50

REGIONAL ACCOUNTS

Structure and Performance of the New York Region's Economy in the Seventies

Prepared by Regina Belz Armstrong

This pioneering study of interlocking accounts for the New York-New Jersey-Connecticut Urban Region traces the flow of funds through income, consumption, and production within a regional economy, and between the Region and the rest of the nation. (A Regional Plan Association Book)

256 pages, tables, charts $18.50

PROPHETS OF PROSPERITY

America's First Political Economists

Paul K. Conkin

The development of a major discipline as seen through the theories of its ablest proponents. In-depth profiles of these pioneers trace the diversity of American economic thought from its European origins through a growing national awareness.

352 pages, index $22.50

MANAGEMENT OR CONTROL?

The Organizational Challenge

Russell Stout, Jr.

"... the best analysis of the confusion which attends the use of management-control systems and in fact the only sustained ... inquiry of both the theory and practice of such systems. ... a fine piece of work." —Martin Landau, University of California, Berkeley

224 pages, bibl., notes, indexes $12.95

ORGANIZATIONS, MANAGEMENT, AND CONTROL

An Annotated Bibliography

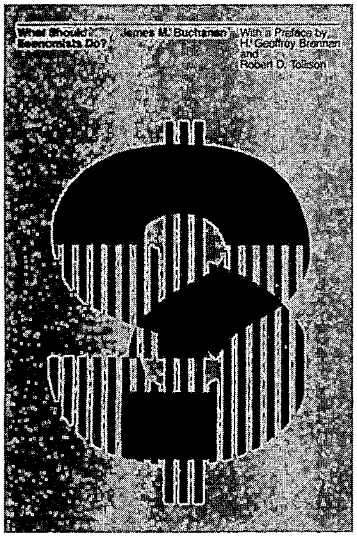
Russell Stout, Jr.

The result of years of scrupulous research, this volume offers the practicing manager and researcher a valuable guide to the literature on organizations and management. Features approximately 1500 entries and draws special attention to outstanding contributions in the field.

256 pages, index $15.00

Available at bookstores or send $1.50 postage and handling for first book, 25¢ for each additional book, to order from publisher.

INDIANA UNIVERSITY PRESS
Tenth and Morton Streets, Bloomington, Indiana 47405


What Should Economists Do? James M. Buchanan With a Preface by H. Geoffrey Brennan and Robert D. Tollison

Liberty*Press*
Liberty*Classics*

What Should Economists Do?

By James M. Buchanan

This collection brings together Buchanan's important essays on method, many of them previously unpublished. Such a volume, note H. Geoffrey Brennan and Robert D. Tollison in their preface, "provides relatively easy access to a group of significant papers on methodology in economics, written by a man whose work has spawned a methodological revolution in the way economists and other scholars think about government and governmental activity."

James M. Buchanan is University Distinguished Professor and General Director of the Center for Study of Public Choice, Virginia Polytechnic Institute and State University. As a founder of the burgeoning subdiscipline of public choice, and as a public finance theorist and father of a modern school of public finance, Buchanan's work has had worldwide recognition. Hardcover $8.00, Paperback $3.50.

We pay postage, but require prepayment, on orders from individuals. Please allow four to six weeks for delivery. To order this book, or for a copy of our catalog, write:

Liberty*Press*/Liberty*Classics*
7440 North Shadeland, Dept. 650
Indianapolis, Indiana 46250

Announcing two New Volumes
in the *INSTITUTE FOR RESEARCH ON POVERTY MONOGRAPH* Series

Microeconomic Simulation Models for Public Policy Analysis

Volume 1/DISTRIBUTIONAL IMPACTS

Edited by ROBERT H. HAVEMAN and KEVIN HOLLENBECK

CONTENTS: EXTENDED MICRODATA MODELS FOR FIRST-ROUND DISTRIBUTIONAL ANALYSIS. *J. J. Minarik*, The MERGE 1973 Data File. *H. Beebout*, Food Stamp Policy Modeling: An Application of MATH. *G. H. Orcutt et al.*, Microanalytic Modeling and the Analysis of Public Transfer Policies. MICRODATA MODELS WITH BEHAVIORAL RESPONSE. *M. Maxfield, Jr.*, Aspects of a Negative Income Tax: Program Cost, Adequacy of Support, and Induced Labor Supply Reduction. *D. Betson et al.*, A Microsimulation Model for Analyzing Alternative Welfare Reform Proposals: An Application to the Program for Better Jobs and Income. *J. A. King*, The Comprehensive Human Resources Data System: A Model for Estimating The Distributional Impacts of Energy Policies. *D. E. Yett et al.*, The HRRC Health Care Sector Simulation Model. References. Index.

1980, 320 pp., $29.50 ISBN: 0-12-333201-X

Volume 2/SECTORAL, REGIONAL, AND GENERAL EQUILIBRIUM MODELS

CONTENTS: MICRODATA MODELS OF THE MACROECONOMY. *R. L. Bennett and B. R. Bergmann*, Policy Explorations with the Transactions Model of the U.S. Economy. *G. Eliasson*, Experiments with Fiscal Policy Parameters on a Micro to Macro Model of the Swedish Economy. *D. Fullerton et al.*, Corporate and Personal Tax Integration in the United States: Some Preliminary Findings. MICRODATA MODELS WITH REGIONAL AND/OR SECTORAL IMPACTS. *R. H. Haveman et al.*, A Microeconomic Simulation Model for Analyzing the Regional and Distributional Effects of Tax-Transfer Policy; An Analysis of the Program for Better Jobs and Income. *L. Ozanne and J. E. Vanski*, Rehabilitating Central-City Housing: Simulations with the Urban Institute Housing Model. *S. P. Dresch and D. A. Updegrove*, IDIOM: A Disaggregated Policy-Impact Model of the U.S. Economy. AN APPRAISAL OF MICRODATA SIMULATION MODELS. *K. J. Arrow*, Microdata Simulation: Current Status, Problems, Prospects. References. Subject Index.

1980, 304 pp., $29.50 ISBN: 0-12-333202-8

Volumes in the *ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS* Series

Studies in Macroeconomic Theory

By EDMUND S. PHELPS

Volume 2/REDISTRIBUTION AND GROWTH

The book is divided into five main sections, each prefaced by a commentary written for this publication. These sections treat issues such as:

- The welfare economics of alternative steady states and their fiscal requirements:
- The types of national investment available to raise productivity, and the nature of their working:
- The alternative notions of socially optimal economic growth, from "fiscal neutrality" to "game equilibrium" to "Rawlsian 'maximin' ":
- The optimal structure of taxation for the present generation under first-best and second-best constraints that protect future generations.

1980, 384 pp., $21.00 ISBN: 0-12-554002-7

Volume 1/EMPLOYMENT AND INFLATION

CONTENTS: Disequilibrium Models of Employment. Macroequilibrium Employment. Welfare Analysis of Anticipated Inflation: Liquidity and Saving. Inflation Planning. Identifying Disturbances of the Price Level. Stabilization Theory. Toward Just Criteria in Stabilization Policy.

1979, 432 pp., $24.50 ISBN: 0-12-554001-9

Texas National Energy Modeling Project

An Experience in Large-Scale Model Transfer and Evaluation

Edited by MILTON L. HOLLOWAY

In this book, the Texas National Energy Modeling Project evaluates the Department of Energy's primary energy forecasting and policy analysis tool—the Midterm Energy Forecasting System (MEFS)—to determine its workability, clarity, validity, and verifiability. This evaluation will provide guidance to users of MEFS concerning the level of confidence one may have in the results of the models used for government energy policy analysis purposes.

1980, 160 pp., $14.50 ISBN: 0-12-352950-6

Continuation Orders authorize us to ship and bill each volume in a series, or "Advances" type publication automatically, immediately upon publication. This order will remain in effect until cancelled. Specify the volume title with which your order is to begin.

Send payment with order and save postage and handling charge.
Prices are subject to change without notice.

ACADEMIC PRESS, INC.
A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003/24-28 OVAL ROAD, LONDON NW1 7DX

November Pt 2
Rational Expectations

JOURNAL OF MONEY, CREDIT AND BANKING

Leading proponents and critics of the theory of rational expectations participated in the American Enterprise Institutes Seminar on Rational Expectations, held on February 1, 1980, in Washington, D.C. Papers and comments from the seminar will be mailed to **JMCB** subscribers as a complimentary, separately bound supplement to the November 1980 issue.

Bennett T. McCallum, Editor • **Robert E. Lucas, Jr.** Methods and Problems in Business Cycle Theory • **Bennett T. McCallum** Rational Expectations and Macroeconomic Stabilization Policy: An Overview • **Robert J. Barro** Federal Deficit Policy and the Effects of Public Debt Shocks • **William Fellner** The Valid Core of Rationality Hypotheses in the Theory of Expectations • Comments: **James Tobin** Are New Classical Models Plausible Enough to Guide Policy? • **Edwin Burmeister** On Some Conceptual Issues in Rational Expectations Modeling • **Arthur M. Okun** Rational-Expectations-with-Misperceptions as a Theory of the Business Cycle • **Phillip Cagan** Reflections on Rational Expectations • **Gottfried Haberler** Critical Notes on Rational Expectations

JMCB is published quarterly. About 900 pages in 1980. Subscriptions, including "Rational Expectations": U.S. $17.50 per year (overseas postage $1.90). "Rational Expectations" only, U.S. $5.00. Send your order to the Ohio State University Press, 2070 Neil Ave., Columbus, OH 43210. 614-422-6930

NEW
PRINCETON
PAPERBACKS

POLITICAL CONTROL OF THE ECONOMY
Edward R. Tufte
"A revealing look into the uses of economic policy for political purposes."
—The Wall Street Journal
"Brilliant." —Business Week
". . . Read this book. It's short, jargon-free, readable, informative and free of ideological axes, which is saying a lot for a book about the 'dismal science.' "
—Mark Green, The Washington Post Book World
$4.95 (Cloth, $10.00)

THEORY OF GAMES AND ECONOMIC BEHAVIOR
John von Neumann and Oskar Morgenstern
"Posterity may regard this book as one of the major scientific achievements of the first half of the twentieth century."
—The Bulletin of the American Mathematical Society
"[Its] main achievement . . . lies, more than in its concrete results, in its having introduced into economics the tools of modern logic and in using them with an astounding power of generalization." —The Journal of Political Economy
$9.95 (Cloth, $30.00)

PRINCETON
UNIVERSITY PRESS
PRINCETON, NEW JERSEY 08540

Modern Theories of Income Distribution

M.C. Howard

A systematic analytical survey of theories of income distribution as they have developed since the 1930's, and particularly since the 1950's, this book elucidates the controversies which have complicated modern income distribution theory, and evaluates the current status of the subject.

1980 ISBN 0-312-54244-5 232 pp., cloth $25.00

The Theory of Economic Growth

An Introduction

Graham Hacche

The most comprehensive and up-to-date treatment of growth theory, this book moves beyond other works on the subject in providing a full account of both neo-Keynesian theory and the neo-classical approach, as well as an extensive discussion of empirical interpretation and application of growth theory.

1980 ISBN 0-312-79775-3 349 pp., cloth $27.50

Economics and Marxism

Volume 1 The Renaissance of the Marxian System

Volume 2 The Dynamics of the Marxian System

Karl Kühne; translated by Robert Shaw

Economics and Marxism offers a comprehensive reassessment of Marx's economics in the light of modern economic theory. In each volume Kühne first discusses the views of academic and Marxist economists then reconstructs Marx's own account. Volume 1 covers value and surplus value, price theory, concentration theory, monopolies and oligopolies, profit and rent theory, wage theory, and monetary theory; Volume 2, international trade, growth business cycles, Marx and Keynes, imperialism and Marxist dynamics.

1980 Vol. 1: 413 pp., cloth Vol. 2: 371 pp., cloth each $27.50
Vol. 1: ISBN 0-312-23436-8 Vol. 2: ISBN 0-312-23437-6

An Annotated Bibliography of Health Economics

Western European Sources

edited by Adrian Griffiths, N. Prescott, R. Rigoni and P. Tacier

The companion volume to *An Annotated Bibliography of Health Economics: English Language Sources*, this book has been clearly organized under the same headings as the previous volume. The informatively annotated entries are listed in chronological order within chapters or subsections, and alphabetically within each year.

November ISBN 0-312-03874-7 about 416 pp., cloth $40.00

An Annotated Bibliography of Health Economics

English Language Sources

edited by A.J. Culyer, J. Wiseman and A. Walker

"An excellent English bibliography covering the relatively new field of health economics..." *Choice*

1977 ISBN 0-312-03873-9 361 pp., cloth $29.95

To order these books for your library at 20% discount, please write:

St. Martin's Press P.O. Box 5352 New York, N.Y. 10163

St. Martin's Press

NORTH-HOLLAND ANNOUNCES

NORTH-HOLLAND PUBLISHING COMPANY
52 Vanderbilt Ave., New York, N.Y. 10017/ P.O. Box 211, 1000 AE Amsterdam, The Netherlands

NEW BOOKS·NEW BOOKS·NEW BOOKS·NEW BOOKS·NEW

Inflation, The Quantity Theory, and Rational Expectations

Studies in Monetary Economics, Vol. 5

By EDUARD JAN BOMHOFF, Erasmus Univ., Rotterdam, The Netherlands

1980. 254 pages.
US $29.25/Dfl. 60.00
ISBN 0-444-85472-X

In this volume a small macro-economic model that can explain the rate of inflation and the rate of econmic growth is developed and tested. Throughout the model much emphasis is placed on the way economic behavior is influenced by expectations, both with respect to the immediate future and regarding long-run developments.

Exhaustible Resources, Optimality, and Trade

Contributions to Economic Analysis, Vol. 127

Edited by MURRAY C. KEMP, Univ. of New South Wales, and NGO VAN LONG, Australian National Univ.

1980. 264 pages.
US $44.00/Dfl. 90.00
ISBN 0-444-85339-1

A collection of essays written between 1974 and 1978 focusing on theoretical issues regarding general equilibrium models accomodating the exhaustibility of resources.
CONTENTS: 1. The Theory of Mining Firm. 2. Aggregative Models of a Closed Economy. 3. International Trade with Exhaustible Resources.

Capital and the Distribution of Labor Earnings

Contributions to Economic Analysis, Vol. 126

By MICHAEL SATTINGER, Miami University of Ohio

1980. 300 pages.
US $46.25/Dfl. 95.00
ISBN 0-444-85385-8

This volume investigates the relation between the distribution of capital among jobs and the distribution of labor earnings among individuals. The major model may be regarded as an extension of David Ricardo's theory of differential rents, A. D. Roy's model of self-selection in labor markets and Jan Tinbergen's theory of the pricing of heterogeneous labor units in a supply and demand multimarket equilibrium.

A Methodological Appraisal of Marxian Economics

Lectures in Economics: Theory, Institutions, Policy, Vol. 3

By MARK BLAUG, Univ. of London, Institute of Education

1980. 96 pages.
US $24.50/Dfl. 50.00 Paperback
ISBN 0-444-85424-X

The aim of the lectures presented here is to appraise Marxian economics as a scientific research programme in the sense of Lakatos, that is, to assess its predictive record for the class of events with which it is concerned.
CONTENTS: 1. Marx and Marxists on the Nature of Profits. 2. Marx and Marxists on the Future of Capitalism. 3. Marx on the History of Economic Thought.

Prospects of Economic Growth

Edited by S. K. KUIPERS and G. J. LANJOUW, Univ. of Groningen, The Netherlands

1980. 295 pages.
US $46.25/Dfl. 95.00
ISBN 0-444-85355-3

This book covers the various aspects of the prospects of economic growth in the Western industrialized countries, and deals with the following 8 major topics: 1. changes in the economic order at the national level and the relation with economic growth, 2. changes in the economic order at the international level and their relation to economic growth, 3. economic growth, technical progress and natural resources, 4. economic growth and environment, 5. prospects of employment growth, 6. shifts in the sectoral pattern of production, 7. the long wave in economic life, and 8. the possibilities of long-run forecasting.

0728 NH

NORTH-HOLLAND ANNOUNCES

NEW BOOKS • NEW BOOKS • NEW BOOKS • NEW BOOKS • NEW

NORTH-HOLLAND PUBLISHING COMPANY
52 Vanderbilt Ave., New York, N.Y. 10017/ P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Economic Dynamics: Methods and Models

Advanced Textbooks in Economics, Vol. 16

By GIANCARLO GANDOLFO, University of Rome, Italy

1980. 560 pages.
US $46.25/Dfl. 95.00
ISBN 0-444-85419-3

This textbook, a major revision of the author's previous book, *Mathematical Methods and Models in Economic Dynamics*, shows how the mathematical methods commonly used in economic dynamics are utilized to build and to analyze dynamic models. The focus is accordingly on methods, and every new mathematical technique introduced is followed by its application to selected models. The material is carefully organized and presented, worked out in great detail, and complemented by many exercises and references. It ranges from an elementary level to rather advanced methods and problems, so that the book is suitable for both undergraduate and graduate use, for beginners and advanced students, as well as for use as a reference work.

Multiple Objective Decision Methods and Regional Planning

Studies in Regional Science and Urban Economics, Vol. 7

By PIETER RIETVELD, Free University, Amsterdam

1980. 340 pages.
US $36.50/Dfl. 75.00
ISBN 0-444-86001-0

This work places multiple objective decision methods in a broad multidisciplinary framework: planning theory, psychology, economics and philosophy. Special attention is paid to the communication aspects of decision-making: the basic question dealt with is how an analyst can provide digestible information to decision-makers about the conflicts and compromises implied by complex decision problems.

New Scarcity and Economic Growth

More Welfare Through Less Production?

By ROEFIE HUETING, Netherlands Central Bureau of Statistics

1980. 284 pages.
US $29.25/Dfl. 60.00
ISBN 0-444-85400-2

The study describes the deterioration of the environment as a rapidly spreading new variety of scarcity. The scarcity is caused by the increasing demands made on the functions (possible uses) of the environment. This work is an important contribution to the knowledge and tools that maybe once will permit decision makers to balance the benefits of immediate economic growth against the decline in national environmental wealth.

In preparation:

Distributed Lags
Problems of Estimation and Formulation
Second Revised Edition
By PHOEBUS J. DHRYMES

Economics of Shortage
By JANOS KORNAI

A Theory of the Multiproduct Firm
By KENNETH LAITINEN

Studies in Economic Theory and Practice
Essays in Honour of Edward Lipinski
Editor: JERZY ŁOS

Labor Unions and the Inequality of Earned Income
By JOHN S. PETTENGILL

The Fix-Point Approach to Interdependent Systems
Editor: HERMAN WOLD

0729 NH

Now fully up to date:
First half of Report 1979 is out.

Bibliographie
der Wirtschaftswissenschaften
(Bibliography of Economics)

(bis Berichtsjahr 1967: Bibliographie der Sozialwissenschaften/
Until 1967: Bibliography of the Social Sciences).

Internationale Dokumentation der Buch- und Zeitschriftenliteratur der Wirtschaftswissenschaften / International Documentation about books and periodicals in the field of Economics.

Redaktion/Editor:
Bibliothek des Instituts für Weltwirtschaft an der Universität Kiel (Zentralbibliothek der Wirtschaftswissenschaften in der Bundesrepublik Deutschland / Central Library of Economics in the Federal Republic of Germany).

Last publication:
Vol. 71, New Series Vol. 30/1979, I. Cloth DM 152,–
This volume will be completed in Autumn 1980.

From the 1962 issue onward the bibliography also contains an additional classification system and indices (subject and regional index) in English. When purchased together with the ‚Jahrbuch für Sozialwissenschaft‘ (yearbook for the Social Sciences) there will be a reduction in price.

Bibliography of Economics Classification System:
General Bibliographies. Commemorative Publications and Reference Books / Economic History / History, Objectives and Methods of Economics / Economic Theory / National Income Accounts. Input-Output / Economic Policy / Economic Reports. Statistics. Empirical Studies / Public Finance / Money and Credit / Economic Ressources. Infrastructure / Population / Labour. Household Income. Property / Social Policy. Public Health. Education, Science and Technology / Regional Development. Housing / Competition. Economic Concentration / Business Administration / Sectors of the Economy / International Economic Relations.

Vandenhoeck & Ruprecht · Göttingen und Zürich

AEA sponsored Group Life Insurance for you and your family– at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA has joined a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by dividend credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 45% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future dividends and credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program **B-5**
1707 L Street, N.W.—Suite 700
Washington, D.C. 20036

Please send me more information about the AEA Life Insurance Plan.

Name ________________________________ Age ______

Address ______________________________________

City ____________________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Managing Editor:
Alan Rufus Waters
Oklahoma State University

Journal of Economic Development

Volume 4 Number 2 December 1979

BOARD OF EDITORS:
Dale W. Adams
Ohio State University
Choong Yong Ahn
Chung-Ang University
Jo Sup Chung
Chung-Ang University
Richard H. Day
University of Southern California
In Kie Kim
Chung-Ang University
Jeffrey B. Nugent
University of Southern California
Attiat F. Ott
Clark University
Noel D. Uri
Department of Energy
Richard Weisskoff
Iowa State University
Jang H. Yoo
Virginia Commonwealth University
Pan A. Yotopoulos
Stanford University

Articles

7 A Demographic, Economic, and Social Summary of Major Non-Oil Developing Countries *David R. Kamerschen*
23 The Relationship Between Operating and Investment Expeditures: A Problem of Planning and Budgeting *David L. Anderson and T. A. Pinfold*
43 On Exchange Rate Policy in a World of Generalized Floating: The Case of Korea *Won-Key Kwon*
61 Export Development, Product Life Cycle and LDCs: The Case of Korea *Jeffrey S. Arpan and D. S. Kim*
89 United States and Japanese Direct Investment in Korea: A Comparative Study *Chung H. Lee*
115 International Trade in Manufactures Among the Developing Countries *Jaleel Ahmad*
129 Measuring Internal and External Costs of Sanitary Landfill Sites in Chicago Metropolitan Area *Eui-Gak Hwang*
141 Emigration and Growth: A Model for Small Developing Economics *Yiu-Kwan Fan*
155 Contributions of the Determinants to the Rate of Change in the Money Stock: The Case of Korea *In Kie Kim*
169 Economic Factors in the Decline of Fertility in Japan *Masanori Hashimoto*

Submission of Manuscripts:
Contributors in the U.S. and other countries should send their manuscripts to:

Jang H. Yoo
Department of Economics
Virginia Commonwealth University
Richmond, Virginia 23284

Contributors in Korea should send their manuscripts to:

In Kie Kim
Department of Economics
Chung-Ang University
Heuksuk-Dong, Kwanak-Ku,
Seoul, Korea

Manuscripts should be written in English, and the JED's writing style will be sent to the authors as soon as the referees accept the contents.

Annual Subscription Price:
US $10.00 per year for individuals and US $20.00 for institutions. Payment should be forwarded to Jang H. Yoo, School of Business, Virginia Commonweath University, Richmond, Virginia 23284. Outside the U.S., send payment to In Kie Kim, Department of Economics, Chung-Ang University, Heuksuk-Dong, Kwanak-Ku, Seoul, Korea.

THE ECONOMICS OF INTERDEPENDENCE

Economic Policy in the Atlantic Community

Published for the Council on Foreign Relations

RICHARD N. COOPER. Now back in print, this volume offers a penetrating analysis of the complex and often contradictory economic currents flowing between the two hemispheres. Richard N. Cooper addresses the central problem facing the world's macroeconomists: how to preserve the manifold benefits accruing from international trade while maintaining sufficient domestic freedom to pursue attractive political and economic objectives. *A Morningside Book.* 316 pages, $20.00 cloth, $6.50 paper

DOMESTIC AND MULTINATIONAL BANKING

The Effect of Monetary Policy

RAE WESTON. This comprehensive book demonstrates the impact of monetary policy on domestic and international banking. The author uses a blend of theory and empirical analysis to examine the progressive development of commercial banking, commencing with a model of a completely unregulated bank and moving to the regulated domestic bank, then to international and multinational banking. 240 pages, $32.50

UNITED STATES TAXATION AND DEVELOPING COUNTRIES

EDITED BY ROBERT HELLAWELL. This book explores United States tax policy as it relates to investment and other operations of multinational corporations in developing countries. In dealing with a wide range of domestic and foreign policy issues, the contributors address the thorny question of whether the U.S. should actively encourage its businesses to invest in developing countries, and whether a neutral U.S. tax policy is possible. *Columbia University Center for Law and Economics.* 448 pages, $30.00

GLOBAL FOOD INTERDEPENDENCE

Challenge to American Foreign Policy

RAYMOND F. HOPKINS AND DONALD J. PUCHALA. Throughout this century the United States has been the world's major exporter of food, influencing the prices, flow of trade, incentives for production, and degree of food security for other countries. *Global Food Interdependence* is a searching examination of United States food policies in the areas of production, trade, and aid—policies which directly affect broader reaches of international affairs. 224 pages, $20.00 cloth, $7.00 paper

Order any of the books advertised above by sending check or money order to **Dept. JN** at the address below, and including $1.30 per order for postage and handling.

COLUMBIA UNIVERSITY PRESS

136 South Broadway, Irvington, New York 10533

Announcing the

Housing Finance Review

a new quarterly journal sponsored by the Federal Home Loan Bank Board
The first issue will appear in January 1981.

Objectives

The **Review** will present the results of research in housing finance and will provide a forum for a continuing dialogue on policy issues related to housing finance. All aspects of housing finance will be considered, regardless of institutional focus.

Authors of high-quality, appropriate manuscripts are invited to submit them for consideration. All papers will be refereed.

Associate Editors

Kenneth R. Biederman
Jack M. Guttentag
Patric H. Hendershott
Dwight M. Jaffee
Michael A. Jessee
Edward J. Kane
Donald M. Kaplan
Marshall A. Kaplan
Sherman J. Maisel
Maurice Mann
Richard G. Marcis
Allan Meltzer
Dale P. Riordan
Kenneth T. Rosen
Craig Swan
Kevin E. Villani
Susan M. Wachter

A free sample issue may be obtained by writing to the Editor.

For complete information, authors' guide, and to submit a manuscript write to:

A. Thomas King, Editor
Housing Finance Review
Federal Home Loan Bank Board
1700 G Street, N. W.
Washington, D. C. 20552

THE ECOLOGY OF HOUSING DESTRUCTION

By Peter D. Salins

An International Center for Economic Policy Studies Book (ICEPS)

"Two hundred thousand apartments have been destroyed within a decade. Another two hundred thousand are or are about to be owned by the city of New York, and thus are in the pipeline for imminent destruction."

Thus, author Peter D. Salins characterizes the consequences of "The Ecology of Housing Destruction," a process whose origins he traces to a set of interlocking public policies and private sector dynamics, many of them unique to New York City.

Other cities are considering the same government intervention measures that New York has employed for the past 30 years. THE ECOLOGY OF HOUSING DESTRUCTION is a timely and illuminating study of the existing policies and regulations which were meant to assist—but have actually destroyed the housing of New York City. 176 pages, $12.50 cloth, $5.95 paper

Other books in the ICEPS series:

The New Protectionism
The Welfare State and International Trade
By Melvyn B. Krauss
119 pages, $10.00 cloth, $4.95 paper

The U.S. Balance of Payments
and the Sinking Dollar
By Wilson Schmidt
157 pages, $10.00 cloth, $4.95 paper

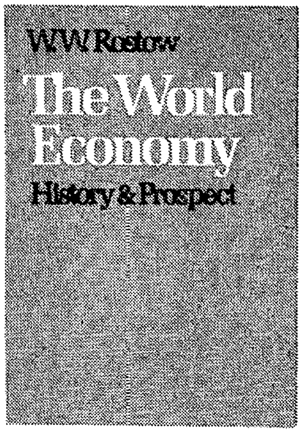
Available at your bookstore, or direct from:

NEW YORK UNIVERSITY PRESS
Washington Square New York, New York 10003

The World Economy

HISTORY AND PROSPECT

By W. W. Rostow



This monumental study is an account of the world economy since the eighteenth century, an analysis and prescription for the future, and a challenge to the reigning neo-Keynesian theories of income determination and growth. No other economic history of this depth and breadth exists. It has become an invaluable reference for economists, economic historians, and other social scientists.

"The scope of the book is astonishing . . . it traces the course of the world economy since the 18th century."
—Lindley H. Clark, Jr., *Wall Street Journal*

". . . the path of evolution from the pre-industrial to the advanced industrial condition is meticulously chronicled."
—Jock Bruce-Gardyne, *Daily Telegraph*

". . . one stands in awe at the breadth of Mr. Rostow's learning, of the enormous amount of labour which he gave to his task. . . ."
—Shepard B. Clough, *Journal of European Economic History*

877 pp., $20.00, paper

TEXAS

University of Texas Press
P.O. BOX 7819 AUSTIN, TEXAS 78712

INTRODUCING
ECONOMICS READING LISTS AND COURSE OUTLINES

Volumes of Current Reading Lists By the World's Leading Economists
Compiled by Edward Tower, Duke University and Simon Fraser University
REVIVE OLD COURSES! CREATE NEW ONES! EXPAND THE POSSIBILITIES FOR INDEPENDENT RESEARCH AND STUDY! THIS WIDE COLLECTION OF FIRST RATE READING LISTS CAN ASSIST EVERY ECONOMIST!

Vol. 1	MICROECONOMICS including cost benefit analysis, game theory, social insurance and the economics of technological change.	145 p.	$11.00
Vol. 2	MACROECONOMICS, MONETARY ECONOMICS, AND MONEY AND BANKING	212 p.	$12.00
Vol. 3	DEVELOPMENT ECONOMICS	180 p.	$11.00
Vol. 4	INDUSTRIAL ORGANIZATION AND REGULATION	154 p.	$11.00
Vol. 5	INTERNATIONAL ECONOMICS	220 p.	$12.00
Vol. 6	PUBLIC FINANCE	131 p.	$10.00
Vol. 7	COMPARATIVE ECONOMIC SYSTEMS	152 p.	$11.00
Vol. 8	ECONOMICS OF LAW, POLITICS AND WAR including public choice, radical economics, altruism, corruption, and the economics of race, class, sex and property rights	138 p.	$11.00
Vol. 9	ECONOMICS OF THE ENVIRONMENT, NATURAL RESOURCES, AGRICULTURE AND ENERGY	134 p.	$10.00
Vol. 10	ECONOMIC HISTORY including America, Europe, The Soviet Union, Britain and East Asia	148 p.	$11.00
Vol. 11	HISTORY OF ECONOMIC THOUGHT	88 p.	$ 8.00
Vol. 12	LABOR ECONOMICS including economics of education and economic-demographic interrelations	109 p.	$10.00
Vol. 13	URBAN AND REGIONAL ECONOMICS including transportation economics	86 p.	$ 8.00
Vol. 14	HEALTH ECONOMICS	86 p.	$ 8.00
Vol. 15	QUANTITATIVE METHODS including econometrics and mathematical economics	160 p.	$11.00

Special price for the complete series $125.00. 30% discount on multiple copies for classroom use. Orders by individuals must be accompanied by payment. Institutions may request billing. Full refund if returned within 30 day examination period. Payment accepted in U.S. Funds or foreign equivalent. Outside North America, add 10% for additional postage. Published July 1980.

Eno River Press
Box 4900, Duke Station
Durham, North Carolina 27706, U.S.A.
(919) 933-7975

SOCIAL CAPITAL AND SOCIALISM

Ilya Stavinsky

This book represents the farthest development of K. Marx's Capital, and the ideas embodied in it are, furthermore, the logical culmination of the classical school of political economy, which numbered among its adherents such eminent economists as V. Petti, Cènè, A. Smith, D. Ricardo, K. Marx and their successors.

If in Capital Marx created a system of categories for the capistalist mode of production based on private capital, in Social Capital and Socialism the author has created a system of categories for the capistalist mode of production which appears in the form of social capital. In other words, he has shown, from the theoretical point of view, the summit of the development of the capitalist mode of production, that is, the transition from the form of private capital "into its opposite, social capital." Furthermore, the theoretical basis of the form of social capital which he has implemented has allowed him to derive quite easily from the latter categories for the socialist mode of production, which were first given by Marx although without strict scientific proof, and he has thus strictly theoretically demonstrated its historical inevitability.

212 pages, $7.00

To order this book, send check or Money Order for $7.00 + $1.00 per order for postage and handling, to the address below:

Ilya Stavinsky
P.O. Box 763, Mad. Sq. Sta.
New York, N.Y. 10159

From THE FREE PRESS

THE WORLDLY ECONOMISTS

ROBERT SOBEL

Sobel examines the personalities, ideas, and impact on national policy of a dozen economists—from Keyserling to Kahn, and Galbraith to Friedman—whose astonishing rise to prominence marks this as The Age of the Economist.

92978 *$15.00*

SOCIAL SECURITY AND PENSIONS IN TRANSITION

Understanding the American Retirement System

BRUNO STEIN

"Useful and provocative on a subject of great current importance and controversy."—Wilbur J. Cohen, former Secretary of Health, Education, and Welfare. "An extremely timely book which is clearly written and very useful....a most welcome contribution."—Victor R. Fuchs, author of *Who Shall Live?*

93085 *$14.95*

DICTIONARY OF BUSINESS AND ECONOMICS

CHRISTINE AMMER & DEAN S. AMMER

More than 3,000 entries provide detailed, jargon-free definitions of every essential term in economic theory and history, accounting, banking and finance, management and other areas. Illustrated with charts, tables, diagrams, and numerous examples.

90059 *$22.50*

MARKETS AND HIERARCHIES:

ANALYSIS AND ANTITRUST IMPLICATIONS

A Study in the Economics of Internal Organization

OLIVER E. WILLIAMSON

"A powerful addition to our knowledge in this area and equally a stimulus to further research." —Kenneth J. Arrow, Harvard University. "Deserves serious attention from anyone who is interested in public policy toward private enterprise." —*Challenge*

93536 *$16.95*

Order from: **THE FREE PRESS**

A Division of Macmillan Publishing Co., Inc.

100D BROWN STREET, RIVERSIDE, NEW JERSEY 08370

THE ECONOMICS INSTITUTE

Boulder, Colorado

offering

Intensive Preparation For
Post-Graduate Programs
In Economics And Business
For Students From Abroad

☐ English ☐ Mathematics ☐ Economic Theory

☐ Statistics ☐ Accounting ☐ Computer Applications

☐ Business Organization and Management

Current program dates, tuition and fee schedule, and further information available from:

The Director
Economics Institute
Campus Box 259
University of Colorado
Boulder, Colorado 80309

Sponsored by the American Economic Association

SOCIETY MEMBERS SAVE US$142.00 ON AN ANNUAL SUBSCRIPTION TO . . .

SOCIAL SCIENCE & MEDICINE

An International Journal

Editor-in-Chief: **Peter J. M. McEwan,**
Aberdeenshire, Scotland

SOCIAL SCIENCE & MEDICINE was established to aid in the universal dissemination of important research and theoretical work in all interrelated areas of socio-behavioural sciences and medicine — including psychiatry and epidemiology.

In 1977 it substantially expanded its coverage and its frequency. It now appears eighteen times a year in four parts:

A. Psychology and sociology (six issues)
B. Medical anthropology (four issues)
C. Medical economics (four issues)
D Medical geography (four issues)

The last three parts were identified as primary areas for **SOCIAL SCIENCE & MEDICINE;** each now has its own Editor. Thus there are three quarterly journals plus the existing six issues with a sociological coverage, all sent to the same subscribers. It is, however, possible to subscribe to one or more individual parts, and the subscription rate for each part is available on request.

Subscription Information
Annual subscription (1980) for members of **AEA**

Parts A-D	US$40.00
Part A	US$25.00
Part B	US$20.00
Part C	US$20.00
Part D	US$20.00

Prices include postage and insurance.
FREE SPECIMEN COPY AVAILABLE ON REQUEST
Sterling prices for journals will be sent to Uk and Eire customers on request. Prices subject to change without notice.

Pergamon Press

Headington Hill Hall, Oxford OX3 0BW, UK
Fairview Park, Elmsford, New York 10523, USA

Some Recent Papers

An unequivocal change of policy: prevention health and medical sociology by **A G DAVIS,** *Cranfield Institute of Technology, UK*

Sexual bias in drug advertisement by **E L THOMPSON,** *Hunter College, New York, USA*

The transcultural view: prerequisite to interethnic communication in medicine by **H H WEIDMAN,** *University of Miami, USA*

Health policies in Africa, Latin America: adopting the primary care approach by **T BOSSERT,** *Harvard School of Public Health, USA*

Spatial factors in mental health in Tasmania by **S C DAVEY,** *University of Tasmania* and **G G GILES**

Mortality migration, income and air pollution: a comparative study by **S R BOZZO** et al., *Bookhaven National Laboratory, New York, USA*

Research on health care and the family: a methodological overview by **T J LITMAN,** *University of Minnesota, USA* and **M VENTERS**

Dimensions of crisis in American health care: toward a general theory of crisis and health care by **J L CRANE,** *University of Notredame* and **S L P LEGEAY,** *West Virginia Institute, USA*

Sociological imperialism and the profession of medicine: a critical examination of the thesis of medical imperialism by **P M STRONG,** *University of Oxford, UK*

The challenge of feeding the people: Chile under Allende — Tanzania under Nyerere by **C SCHUFTAN,** *Tulane, USA*

Changes in Cuban health care: technological pessimism by **S GUTTMACHER,** *Columbia, USA* and **R DANIELSON**

An experimental evaluation of crisis intervention by **S BORDOW,** *Australia* and **D PORRITT,** *University of New South Wales, Australia*

282 3/80/9D/24

ALLIED SOCIAL SCIENCE ASSOCIATIONS

December 28-30, 1980 ● DALLAS, TX

The Allied Social Science Associations will provide a professional placement service, at no charge, in **DECEMBER, 1980.** This service will be located in the Dallas Hilton Hotel in Dallas, Texas. Operating hours will be December 28th, 10:00 am - 5:00 pm, December 29th, 9:00 am - 5:00 pm and December 30th, 9:00 am - 2:00 pm. **THERE WILL NOT BE A PROFESSIONAL PLACEMENT SERVICE DURING THE SEPTEMBER MEETINGS.**

FORM REQUEST

For your convenience the National Registry for Economists is offering a preregistration service. Requests for placement forms will be processed between September 1 and November 15 only. Listing your application or vacancies prior to the meeting will expedite service at the placement center. Employment applications and listings of available positions will also be accepted from those who cannot attend the meetings and will be made available for review by the participants.

NOTE: **Requests for forms must be in the Registry office no later than November 15, 1980 and the completed forms must be returned to the National Registry Office no later than December 1.** After the December 1 deadline, completed forms must be submitted in person during the meetings.

HOUSING

A block of rooms has been reserved at the Dallas Hilton and the Sheraton Hotels in Dallas. \$26/singles; \$30/doubles and twins; \$50/one bedroom suites; and \$75/two bedroom suites. A hotel reservation card will be mailed to you immediately upon receipt of your request for a form; please return it as early as possible.

NO FORMS WILL BE MAILED TO YOU PRIOR TO SEPTEMBER 1.

- -

Janet Aschenbrenner
National Registry for Economists
Illinois Job Service
40 West Adams Street
Chicago, Illinois 60603

Name____

____
(Employers list name of Organization or Institution)

Address______________________________________

____
(Applicants - please use home address only)

Check Appropriate Block:

☐ Candidate Application Form (Only one per applicant is required)

☐ Employer Position Listing Form - Indicate number of forms required________ (Usually one per vacancy)

The American Economic Review

ARTICLES

M. FELDSTEIN Inflation and the Stock Market

L. W. LEE A Theory of Just Regulation

E. MANSFIELD Basic Research and Productivity Increase in Manufacturing

R. E. VERRECCHIA
Consensus Beliefs, Information Acquisition, and Market Information Efficiency

G. A. AKERLOF AND B. G. M. MAIN
Unemployment Spells and Unemployment Experience

J. V. HENDERSON
Community Development: The Effects of Growth and Uncertainty

D. O. PARSONS Racial Trends in Male Labor Force Participation

C. MENEZES, C. GEISS, AND J. TRESSLER
Increasing Downside Risk

J. S. CHIPMAN AND J. C. MOORE
Compensating Variation, Consumer's Surplus, and Welfare

P. KRUGMAN Scale Economies, Product Differentiation, and the Pattern of Trade

R. DORNBUSCH AND S. FISCHER
Exchange Rates and the Current Account

J. A. HANSON The Short-Run Relation between Growth and Inflation in Latin America

G. V. JUMP Interest Rates, Inflation Expectations, and Spurious Elements in Measured Real Income and Saving

SHORTER PAPERS: R. E. Lucas, Jr.; K. Pertz; J. H. McCulloch; M. D. Levi and J. H. Makin; E. O. Olsen; G. Brennan and C. Walsh; A. H. Barnett; A. P. Hurter, Jr., J. S. Martinich, and E. R. Venta; D. R. Capozza and R. Van Order; E. Silberberg; V. A. Aivazian and J. L. Callen; P. H. Greenwood and C. A. Ingene; L. Eeckhoudt and P. Hansen; B. B. Gibson; R. Dardis; S. Perrakis; F. D. Arditti and H. Levy; B. L. Benson.

DECEMBER 1980

THE AMERICAN ECONOMIC ASSOCIATION

- Published at George Banta Co., Inc. Menasha, Wisconsin. The publication number is ISSN 0002-8282.

- *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Effective April 1, 1980, Association membership dues for 1980, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$28.75 for regular members with rank of assistant professor or lower, or with annual income of $13,800 or less;

$34.50 for regular members with rank of associate professor, or with annual income of $13,800 to $23,000;

$40.25 for regular members with rank of full professor, or with annual income above $23,000;

$14.25 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $43.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $4.75 to cover extra postage.

- Correspondence relating to the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

- Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212.

Founded in 1885

Officers

President

MOSES ABRAMOVITZ
Stanford University

President-Elect

WILLIAM J. BAUMOL
Princeton University and New York University

Vice Presidents

CARL F. CHRIST
Johns Hopkins University
H. GREGG LEWIS
Duke University

Secretary

C. ELTON HINSHAW
Vanderbilt University

Treasurer

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

ROBERT W. CLOWER
University of California-Los Angeles
MARINA v. N. WHITMAN
General Motors Corporation
HENRY J. AARON
The Brookings Institution
ZVI GRILICHES
Harvard University
MARTIN S. FELDSTEIN
National Bureau of Economic Research, Inc. and Harvard University
ROBERT E. LUCAS, JR.
University of Chicago

Ex Officio Members

TJALLING KOOPMANS
Yale University
ROBERT M. SOLOW
Massachusetts Institute of Technology

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

ALBERT ANDO
PRANAB K. BARDHAN
PETER A. DIAMOND
W. ERWIN DIEWERT
RUDIGER DORNBUSCH
HERSCHEL I. GROSSMAN
PETER W. HOWITT
ANNE O. KRUEGER
WILLIAM H. OAKLAND
MICHAEL PARKIN
ROY RADNER
RICHARD W. ROLL
NANCY L. SCHWARTZ
JAMES P. SMITH
A. MICHAEL SPENCE
WILLIAM S. VICKREY
ROBERT D. WILLIG
S. Y. WU

- Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to George H. Borts, Managing Editor, Box Q, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

- No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

- Copyright American Economic Association 1980. All rights reserved.

December 1980

VOLUME 70, NUMBER 5

Articles

Inflation and the Stock Market *Martin Feldstein* 839

A Theory of Just Regulation *Li Way Lee* 848

Basic Research and Producitivity Increase in Manufacturing *Edwin Mansfield* 863

Consensus Beliefs, Information Acquisition, and Market Information Efficiency *Robert E. Verrecchia* 874

Unemployment Spells and Unemployment Experience *George A. Akerlof and Brian G. M. Main* 885

Community Development: The Effects of Growth and Uncertainty *J. Vernon Henderson* 894

Racial Trends in Male Labor Force Participation *Donald O. Parsons* 911

Increasing Downside Risk *C. Menezes, C. Geiss, and J. Tressler* 921

Compensating Variation, Consumer's Surplus, and Welfare *John S. Chipman and James C. Moore* 933

Scale Economies, Product Differentiation, and the Pattern of Trade *Paul Krugman* 950

Exchange Rates and the Current Account *Rudiger Dornbusch and Stanley Fischer* 960

The Short-Run Relation between Growth and Inflation in Latin America: A Quasi-Rational or Consistent Expectations Approach *James A. Hanson* 972

Interest Rates, Inflation Expectations, and Spurious Elements in Measured Real Income and Saving *Gregory V. Jump* 990

Shorter Papers

Two Illustrations of the Quantity Theory of Money *Robert E. Lucas, Jr.* 1005

Reswitching, Wicksell Effects, and the Neoclassical Production Function: Note *Klaus Pertz* 1015

The Ban on Indexed Bonds, 1933–77 *J. Huston McCulloch* 1018

Inflation Uncertainty and the Phillips Curve: Some Empirical Evidence *Maurice D. Levi and John H. Makin* 1022

Pareto-Desirable Redistribution in Kind:

- Comment *Edgar O. Olsen* 1028
- Reply *Geoffrey Brennan and Cliff Walsh* 1032

The Pigouvian Tax Rule under Monopoly *A. H. Barnett* 1037

A Note on the Separability of Production and Location *Arthur P. Hurter, Jr., Joseph S. Martinich, and Enrique R. Venta* 1042

Unique Equilibria, Pure Profits, and Efficiency in Location Models *Dennis R. Capozza and Robert Van Order* 1046

Harold Hotelling and Marginal Cost Pricing *Eugene Silberberg* 1054

Uncertain Externalities, Liability Rules, and Resource Allocation:

- Comment *Varouj A. Aivazian and Jeffrey L. Callen* 1058
- Reply *Peter H. Greenwood and Charles A. Ingene* 1060

Minimum and Maximum Prices, Uncertainty, and the Theory of the Competitive Firm *Louis Eeckhoudt and Pierre Hansen* 1064

Estimating Demand Elasticities for Public Goods from Survey Data *Betty Blecha Gibson* 1069

The Value of a Life: New Evidence from the Marketplace *Rachel Dardis* 1077

Factor-Price Uncertainty with Variable Proportions: Note *Stylianos Perrakis* 1083

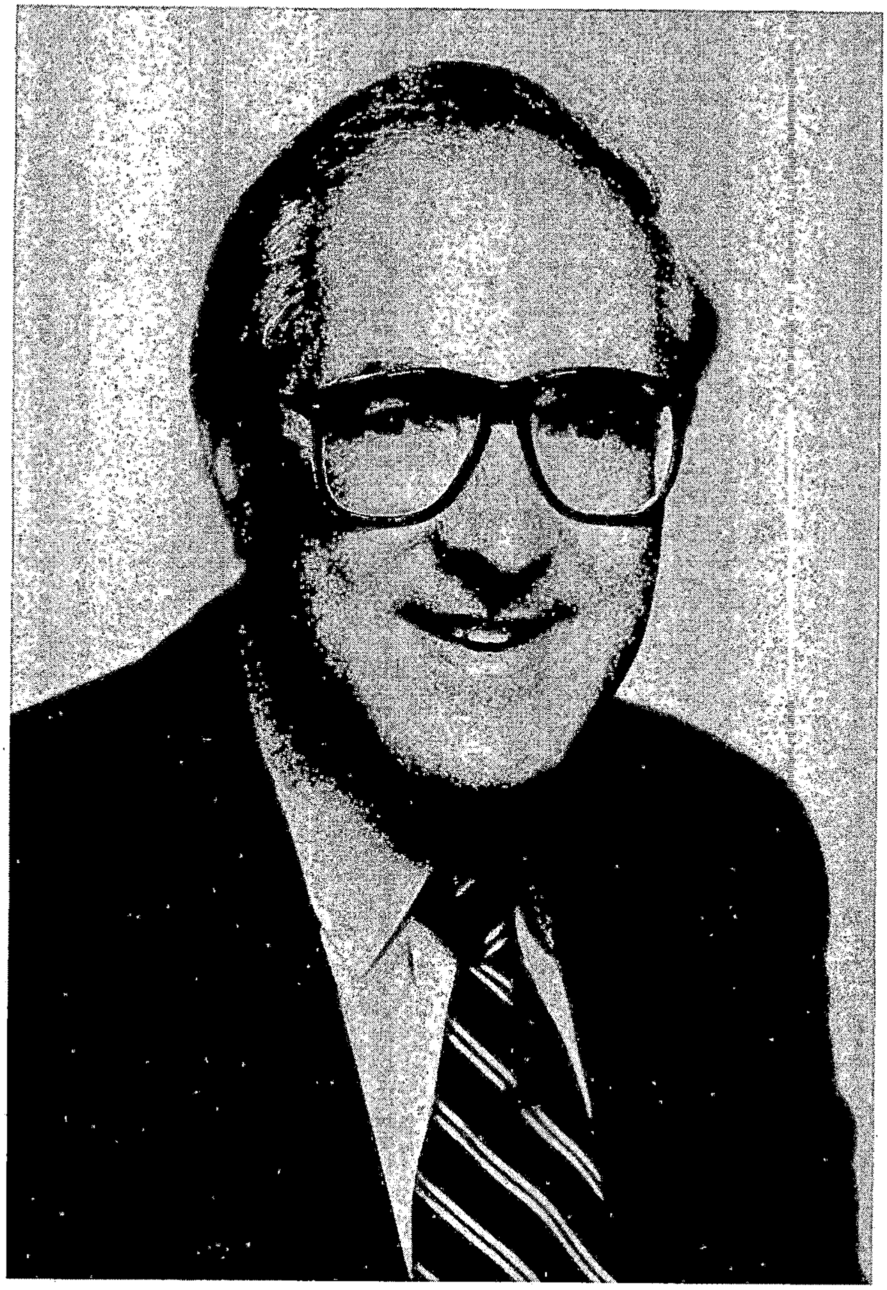
A Model of the Parallel Team Strategy in Product Development *Fred D. Arditti and Haim Levy* 1089

Löschian Competition under Alternative Demand Conditions *Bruce L. Benson* 1098

Notes 1106

Titles of Doctoral Dissertations 1113

GEORGE H. BORTS
Managing Editor of the American Economic Review
1969–1980



George H. Borts

Inflation and the Stock Market

By MARTIN FELDSTEIN*

This paper discusses a crucial cause of the failure of share prices to rise during a decade of substantial inflation. Indeed, the share value per dollar of pretax earnings actually fell from 10.82 in 1967 to 6.65 in 1976.[1] The analysis here indicates that this inverse relation between higher inflation and lower share prices during the past decade was not due to chance or to other unrelated economic events.[2] On the contrary, an important adverse effect of increased inflation on share prices results from basic features of the current *U.S.* tax laws, particularly historic cost depreciation and the taxation of nominal capital gains.[3]

This analysis shows that in order to understand the structural relation between inflation and share prices, it is crucial to distinguish between the effect of a *high* constant rate of inflation and the effect of an *increase* in the rate of inflation expected for the future. When the steady-state rate of inflation is higher, share prices increase at a faster rate. More specifically, when the inflation rate is steady, share prices rise in proportion to the price level to maintain a constant ratio of share prices to real earnings. In contrast, an *increase* in the expected future rate of inflation causes a concurrent fall in the ratio of share prices to current earnings. Although share prices then rise from this lower level at the higher rate of inflation, the ratio of share prices to real earnings is permanently lower. This permanent reduction in the price-earnings ratio occurs because, under prevailing tax rules, inflation raises the effective tax rate on corporate-source income.

This process is illustrated in Figure 1. The top part of the figure shows the inflation rate. Until time t_0, the inflation rate is constant at π_0; it then rises to a higher steady-state level, π_1. This increase is immediately and correctly perceived. The middle part of the figure shows that until t_0 the price per share rises at a constant rate equal to the rate of growth of *nominal* earnings per share. At time t_0, the share price drops to a level consistent with the higher rate of inflation and then grows at the new higher rate of growth of earnings per share. Finally, the lower part of the figure shows that the price-earnings ratio falls when the inflation rate increases but remains constant as long as the inflation rate is constant.

The starting point for this analysis is the way in which inflation raises the effective tax rate on corporate-source income. This is in sharp contrast to the common popular argument that share prices are depressed

---

*President, the National Bureau of Economic Research (NBER), and professor of economics, Harvard University. This study is part of the NBER program of research on Business Taxation and Finance. I am grateful to participants in the NBER Summer Institute for helpful comments, and to the NBER and the National Science Foundation for financial support. The views expressed in this paper are my own and not those of the NBER or Harvard University.

[1] These price-earnings ratios are based on earnings net of real depreciation and with the inventory valuation adjustment. See my paper with Lawrence Summers (1979) for the description of the method by which these pretax price-earnings ratios are constructed. The traditional Standard and Poor's posttax price-earnings ratio, based on book profits (including inventory profits) fell from 17.45 to 9.02. An alternate measure of share-price performance, the ratio of the share price to the underlying real capital at replacement cost, fell from 1.214 in 1967 to 0.788 in 1977. For further evidence of the adverse effect of inflation on real share prices, see Charles Nelson and John Lintner (1973, 1975).

[2] The analysis also shows that it is unnecessary to invoke a theory of systematic error of the type developed by Franco Modigliani and Richard Cohn.

[3] I emphasize that these are tax rules in the United States. Share prices in other countries that do not tax capital gains and that permit extremely rapid tax depreciation of investments may respond very differently to inflation. The relation between share prices and inflation in other countries is therefore of little relevance to the United States without a careful analysis of local tax rules.

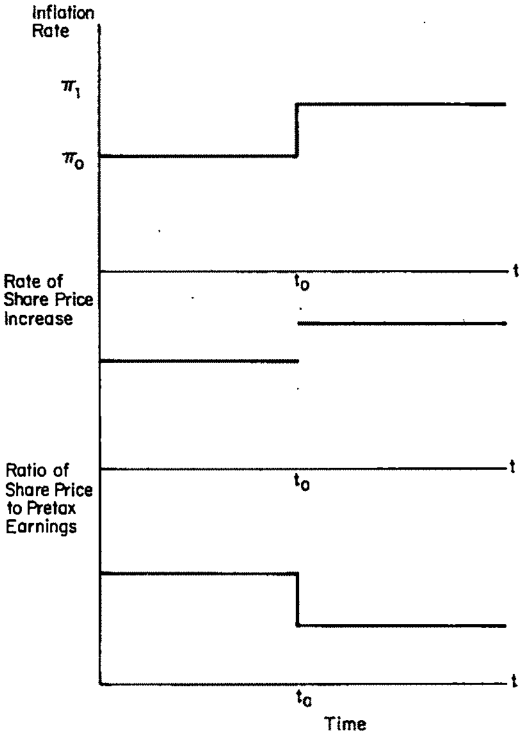


FIGURE 1

because inflation raises the rate of interest that can be earned by investing in bonds. It is clear that this argument should be rejected since the higher *nominal* rate of interest generally corresponds to an unchanged *real* rate of interest.[4] Indeed, since the entire nominal interest is subject to the personal income tax, the real rate of interest net of the personal income tax actually falls. The analysis of Section I shows that, with existing tax rules, inflation is likely to depress the real net rate of interest by less than it lowers the real net return to equity investment. The simple valuation model that calculates the share value by discounting at the real net rate of interest leads to the conclusion that, with current tax laws, an increase in inflation reduces the price that individuals are willing to pay for shares.

Although this discounted earnings model of share valuation is a useful heuristic device, it has the serious deficiency that it implies that individuals in different tax situations would place different reservation values on the same shares. It is therefore inconsistent with the observation that the same stocks are held by individuals who face very different tax rates. The main analysis of this paper therefore uses a more general stock valuation model to derive the assets demanded by investors in different tax situations—and then calculates the share value that achieves a market equilibrium.

Numerical calculations with the market equilibrium model show how inflation can substantially depress the equilibrium share value because of our current tax rules. The model is however very simple. It should be regarded as an aid to understanding and not as a device for making precise calculations.

A complete analysis of the effect of inflation on share prices requires considering a wider range of alternative investments and incorporating the possibility that perceived risk varies with inflation. Explaining the historical experience also requires a model of the process by which both expectations and the capital stock adjust through time. The goal of this paper is more modest: an examination of the way that taxes and inflation interact in affecting equilibrium share prices. The final section of the paper discusses some of the ways in which this analysis could be extended.

I. The Effect of Inflation on the Demand Price of Shares

Consider first an economy in which there is no inflation. Each share of stock represents the ownership claim to a single unit of capital and the net earnings that it produces. There is no corporate debt and all earnings are paid out as dividends. The marginal product of capital (net of depreciation), ρ, is subject to a corporate income tax at rate τ.[5]

---

[4] The conclusion that inflation raises the nominal interest rate while leaving the real rate unchanged has been supported by evidence since Irving Fisher's classic study. For more recent evidence, see my paper with Otto Eckstein, William Yohe and Denis Karnosky, and my paper with Summers (1978).

[5] Since each share of stock represents one unit of capital, the marginal product of capital is also the pretax earnings per share.

The earnings per share that are distributed to the individual investor are therefore $(1-\tau)\rho$. Since there are no retained earnings, the earnings per share do not grow over time and there is no change in the value per share. The individual pays personal tax at rate θ on the earnings that he receives. The individual's net earnings per share are thus $(1-\theta)(1-\tau)\rho$.

A simple model of share valuation implies that the price that the individual would be willing to pay per share (q) would make the net earnings per dollar of equity equal to the net interest that he would receive per dollar invested in government bonds, $(1-\theta)r$.[6] More realistically, individuals may require a higher yield on the riskier equity investment; if this risk differential is denoted by δ, the investor's indifference condition becomes

$$\frac{(1-\theta)(1-\tau)\rho}{q}=(1-\theta)r+\delta \tag{1}$$

The individual's demand price per share is thus

$$q=\frac{(1-\theta)(1-\tau)\rho}{(1-\theta)r+\delta} \tag{2}$$

What happens when the rate of inflation increases from zero to a positive rate π? For simplicity, the analysis will assume an instantaneous and unanticipated increase to π which is then expected to persist forever. To evaluate the new demand price per share, it is necessary to recalculate both the net earnings per share and the real net rate of interest.

Under *U.S.* tax law, taxable profits are calculated by subtracting a value for depreciation from other net operating income. This value of depreciation is based on the original or "historic" cost of the asset rather than on its current value. When prices rise, this historic-cost method of depreciation causes the real value of depreciation to fall and real taxable profits to be increased.[7] As a result, real profits net of the corporate income tax vary inversely with inflation.[8] A linear approximation that each percentage point of inflation reduces net corporate profits per unit of capital by λ implies that net corporate earnings per share of capital are $(1-\tau)\rho-\lambda\pi$.[9] After personal income tax at rate θ, the individual receives $(1-\theta)[(1-\tau)\rho-\lambda\pi]$.[10]

Inflation reduces these net earnings even further by imposing an additional tax on nominal capital gains. More specifically, even though the real share price remains constant at the new equilibrium value q, inflation causes nominal capital gains at the rate of πq.[11] Capital gains are taxed at a lower rate than ordinary income and only when the stock is sold; the equivalent tax rate on accrued capital gains will be denoted c.[12] The extra burden caused by taxing nominal capital gains is thus $c\pi q$. The real net earnings per share are therefore $(1-\theta)[(1-\tau)\rho-\lambda\pi]-c\pi q$. Note that a small increase in the rate of inflation calculated at $\pi=0$ reduces these real net earnings by $(1-\theta)\lambda+cq$.[13]

---

[6] See Mervyn King. He also treats the more general case in which retained earnings cause share prices to rise.

[7] When there is no inflation, the various methods of "accelerated depreciation" that are allowed for tax purposes may cause tax depreciation to exceed the economic depreciation for some assets. This is subsumed in the effective corporate tax rate τ. Accelerated depreciation does not change the conclusion that inflation reduces the real value of depreciation.

[8] For a more complete discussion of this see my paper with Summers (1979) and my paper with Jerry Green and Eytan Sheshinski, especially the appendix by Alan Auerbach. Hai Hong, Brian Motley, and Richard Van Horne and William Glassmire discuss the effects of historic-cost depreciation and the implication for the effect of inflation on share values; they assume a single investor whose discount rate is unchanged by inflation.

[9] It can be shown that with an exponental depreciation rate of 15 percent and a growth rate of 3 percent, a 7 percent inflation rate reduces net profits per unit of capital by 0.021; this implies $\lambda=0.30$. Recall that each share of stock represents a claim to one unit of capital.

[10] This assumes that inflation does not affect the pretax profitability of capital. The calculation also ignores the transitional effect of a lower present value of the future depreciation allowable on past investments.

[11] These nominal capital gains are stated in constant dollars. If the price level at time t is $e^{\pi t}$, the nominal capital gain at that time is $\pi q e^{\pi t}$.

[12] That is, c is the accrual rate of capital gains taxation equivalent to the present value of the tax that will be paid in the future when the stock is sold.

The effect of inflation on the real net rate of interest, $(1-\theta)r-\pi$, depends on the response of the nominal interest rate to the rate of inflation. As noted above, the *U.S.* experience has been that $dr/d\pi=1$. Thus $d[(1-\theta)r-\pi]/d\pi=-\theta$.

For reasonable values of the tax parameters, the decrease in net earnings per dollar of equity $[(1-\theta)\lambda+cq)/q]$ exceeds the decrease in the real net interest yield on bonds $[\theta]$. For example, with a personal tax rate of $\theta=0.3$, a depreciation effect of $\lambda=0.30$, an effective capital gains tax rate of $c=0.15$, and an initial share value per unit of capital of $q=1$, each 1 percent of inflation reduces the real net yield on equity by 0.36 percent and reduces the real net yield on debt by 0.30. If the risk premium (δ) is unchanged, this implies that the share price calculated as the discounted value of earnings per share will fall.

More specifically, the simple valuation model that calculates the individual's demand price per share by equating the real net yield per dollar of equity to the sum of the real net interest rate and the risk premium implies

$$\frac{(1-\theta)[(1-\tau)\rho-\lambda\pi]}{q}\equiv c\pi = (1-\theta)r-\pi+\delta \tag{3}$$

or

$$q=\frac{(1-\theta)[(1-\tau)\rho-\lambda\pi]}{(1-\theta)r-(1-c)\pi+\delta} \tag{4}$$

Differentiating q with respect to π with the condition that $dr/d\pi=1$ implies

$$\frac{dq}{d\pi}=\frac{-(1-\theta)\lambda+q(\theta-c)}{(1-\theta)r-(1-c)\pi+\delta} \tag{5}$$

Since the denominator is positive,[14] $dq/d\pi$ is negative if

$$q(\theta-c)<(1-\theta)\lambda \tag{6}$$

Several things about this condition should be noted. First, realistic values of the tax parameters satisfy the inequality and imply $dq/d\pi<0$. The example of $\theta=0.3$, $c=0.15$, and $\lambda=0.30$ implies $dq/d\pi<0$ even at $q=1$. Second, the inequality is satisfied more easily for investors with low individual tax rates. In the important extreme case of a tax-exempt institution, $\theta=c=0$ and the inequality is satisfied for any value of $\lambda>0$. Finally, for some individuals with high tax rates, the inequality will not be satisfied and $dq/d\pi>0$.[15]

This diversity of responses of q to the rate of inflation reinforces the implication of equation (4) that the demand price per share differs among investors according to their tax situation. An analysis of the effect of inflation on the market price of stock requires a portfolio model of investor equilibrium. In such a model, the risk premium (δ) is both implicit and endogenous. The risk differential changes as the investor reallocates his portfolio until a market equilibrium is achieved in which the same market value of stock is consistent with each investor's own portfolio equilibrium. The specification of such a model is the subject of Section II. Section III then analyzes how the equilibrium responds to a change in the rate of inflation.

II. A Market Equilibrium Model of Share Valuation

The market equilibrium model builds directly on the analysis of the previous section. The economy is assumed to have two assets (risky equity shares and riskless government bonds) and two types of portfolio investors (tax-exempt institutions and taxable individuals). The analysis begins by specifying the investors' portfolio equilibrium equations. When these are combined with the asset supply constraint, they implicitly define the market value per share of equity. This equilibrium model is then used in Section III to calculate the effect of changing from an equilibrium with a zero rate of inflation to a new equilibrium with a positive constant rate of inflation.

---

[13] This is also true for $\pi>0$ if the change in q is ignored.

[14] This is a necessary condition for a finite value of q.

[15] Recall that these calculations all assume a firm with no debt finance and no retained earnings.

The household's investment problem is to divide its initial wealth between bonds and stocks. Equation (1) showed that, in the absence of inflation, the portfolio equilibrium of the households can be written

$$\frac{(1-\theta)(1-\tau)\rho}{q_0}=(1-\theta)r_0+\delta_{h0} \tag{7}$$

where the share price and interest rate carry a subscript zero to distinguish these initial pre-inflation values from the values of these variables when there is inflation. The risk premium δ_{h0} has subscripts to indicate that it refers to the household in the initial equilibrium.

The risk premium that a household requires to hold a marginal share of equity should be an increasing function of the amount of risk that the household is already bearing. More explicitly, I shall assume that δ_{h0} is proportional to the standard deviation of the return on the household's portfolio. The source of this uncertainty is the variability of the pretax equity return; the variance of ρ will be written σ_p^2. The after-tax variance per dollar of equity investment is thus $(1-\theta)^2(1-\tau)^2\sigma_p^2/q_0^2$. If the household has s_{h0} shares in the initial equilibrium, the dollar value of its equity investment is $s_{h0}q_0$. Since bonds are riskless, the variance of the return on the household portfolio is $s_{h0}^2(1-\theta)^2(1-\tau)^2\sigma_p^2$. If the risk premium is proportional to the standard deviation of the portfolio return,

$$\delta_{h0}=\delta_h s_{h0}(1-\theta)(1-\tau)\sigma_p \tag{8}$$

where δ_h is a constant.[16]

Substituting (8) into (7) and rearranging terms yields the share price that is consistent with the household's chosen share ownership in the absence of inflation:

$$q_0=\frac{(1-\theta)(1-\tau)\rho}{(1-\theta)r_0+\delta_h s_{h0}(1-\theta)(1-\tau)\sigma_p} \tag{9}$$

Note that this household demand price for shares varies inversely with the quantity of shares that it holds.

For tax-exempt institutions, the relevant value of θ is zero. The portfolio equilibrium of the institution can therefore be written

$$\frac{(1-\tau)\rho}{q_0}=r_0+\delta_{i0} \tag{10}$$

where the institution's risk premium (indicated by subscript i) satisfies

$$\delta_{i0}=\delta_i s_{i0}(1-\tau)\sigma_p \tag{11}$$

Combining these two equations yields the institution's demand price per share:[17]

$$q_0=\frac{(1-\tau)\rho}{r_0+\delta_i s_{i0}(1-\tau)\sigma_p} \tag{12}$$

The total number of shares outstanding, $\bar{s}$, constrains the combined holdings of the institution and the household investors:

$$\bar{s}=s_{i0}+s_{h0} \tag{13}$$

This supply constraint and the two demand equations (9) and (12) are sufficient to determine the equilibrium share price and the allocation of the shares between the two types of investors. The nature of this solution is illustrated in Figure 2. The equilibrium relation of equation (9) is drawn as the household's demand curve for shares, $s_{h0}(q_0)$. Similarly (12) is drawn as the institution's demand curve $s_{i0}(q_0)$. These are

---

[16] Note that σ_p is a standard deviation of a rate of return and is therefore in the same units as the rate of return, i.e., percent per year. The coefficient δ_h is therefore a unit-file number.

[17] Note that, since $(1-\theta)$ can be eliminated from equation (9), the household and institution demand price equations differ in the absence of inflation only because of differences in δ_i.

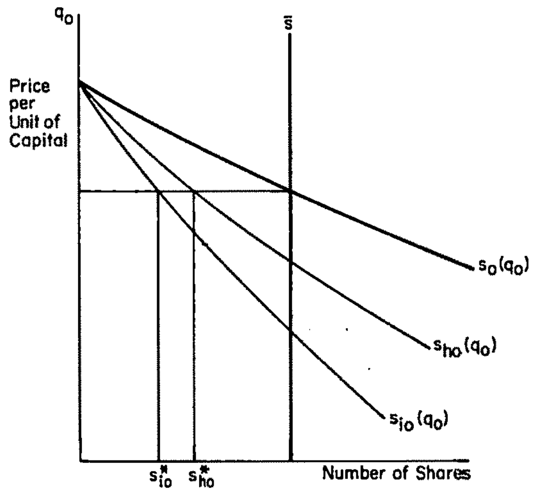


Figure 2

added horizontally to get total share demand as a function of share price $s_c(q_0)$. The intersection of this market demand curve with the supply constraint line $\bar{s}$ determines the equilibrium price q_0^* and the corresponding share holdings (s^*_{h0} and s^*_{i0}).

To use this model to study the impact of inflation on share prices, it is necessary to evaluate the parameters $\delta_h\sigma_p$ and $\delta_i\sigma_p$. Although these parameters cannot be observed, their values can be inferred from the equilibrium conditions in the absence of inflation. Thus equation (9) implies

$$\delta_h\sigma_p = \frac{(1-\tau)\rho - q_0 r_0}{q_0(1-\tau)s_{h0}} \tag{14}$$

and (12) implies

$$\delta_i\sigma_p = \frac{(1-\tau)\rho - q_0 r_0}{q_0(1-\tau)s_{i0}} \tag{15}$$

For these illustrative calculations, the pretax return to capital will be taken to be $\rho = 0.11$, the effective corporate tax rate $\tau = 0.4$, and the interest rate in the absence of inflation $r_0 = 0.03$. In 1970, households held \$700 billion of equities while institutions (including private pension plans, foundations, educational endowments, and insurance companies) held \$135 billion; I will therefore set $q_0 s_{h0} = 700$ and $q_0 s_{i0} = 135$. Together these assumptions imply

$$\delta_h\sigma_p = (0.157 - .071 q_0)10^{-3} \tag{16}$$

$$\delta_i\sigma_p = (0.815 - .370 q_0)10^{-3} \tag{17}$$

It is common to assume that the corporate capital stock (and therefore ρ and $\bar{s}$) will adjust until in equilibrium q equals one. An alternative view is that, with the corporate financial behavior that is optimal under existing tax laws, this arbitrage will not be fully achieved and the equilibrium value of q_0 will be less than one.[18] Section III shows how, under either assumption, the introduction of a moderate rate of inflation can cause a substantial fall in the share value.

III. Inflation and the Market Equilibrium Share Value

This section examines the effect of an unanticipated increase in the steady-state rate of inflation. The analysis assumes that the corporate capital stock remains constant; this implies that the total number of shares ($\bar{s}$) and the average pretax profitability (ρ) remain constant.[19]

Inflation changes the net yields on stocks and bonds in the way described in Section I. For households, the real net yield on equity becomes $(1-\theta)[(1-\tau)\rho - \lambda\pi]/q_1 - c\pi$ and the real net return on bonds becomes $(1-\theta)r_1 - \pi$. Since the nominal interest rises by the rate of inflation,[20] $r_1=r_0+\pi$ and the real net return on bonds is $(1-\theta)r_0-\theta\pi$. The new portfolio equilibrium therefore satisfies

---

[18] See Auerbach, David Bradford, and King for statements of this view.

[19] The assumption of a fixed corporate capital stock causes the calculation to overstate the change in the share price. If q falls, capital will leave the corporate sector, raising ρ and thereby q. Since this would be anticipated by investors, the immediate fall would be less than that calculated here. A satisfactory solution to this problem requires a dynamic model with endogenous corporate investment decisions.

$$(18)\qquad \frac{(1-\theta)[(1-\tau)\rho-\lambda\pi]}{q_1}-c\pi =(1-\theta)r_0-\theta\pi+\delta_{1h}$$

where

$$(19)\qquad \delta_{1h}=\delta_h s_{h1}(1-\theta)(1-\tau)\sigma_p$$

The household's demand price for shares therefore satisfies

$$(20)\qquad q_1=\frac{(1-\theta)[(1-\tau)\rho-\lambda\pi]}{(1-\theta)r_0-(\theta-c)\pi+\delta_h s_{h1}(1-\theta)(1-\tau)\sigma_p}$$

Similarly, the institution's demand price for shares (with $\theta=c=0$) can be written as

$$(21)\qquad q_1=\frac{(1-\tau)\rho-\lambda\pi}{r_0+\delta_i s_{i1}(1-\tau)\sigma_p}$$

Since the number of shares has not changed, it is still true that

$$(22)\qquad \bar{s}=s_{h1}+s_{i1}$$

These three equations determine the new equilibrium share price and the corresponding allocation of shares.

Before calculating the new equilibrium explicitly, it is useful to discuss the change with the help of a diagram. Figure 3 combines the no-inflation demand equations originally shown in Figure 2 with the corresponding demand equations of (20) and (21) in the presence of inflation. The dashed lines ($s_{h0}(q_0)$ and $s_{i0}(q_0)$) show the no-inflation share demands and s_0, the horizontal sum of these demand curves, gives the market demand. Comparing equations (21) and (12) shows that the institutions' demand price is lower at every value of s_{i1} but also tends to zero as s_{i1} tends to infinity; the curve $s_{i1}(q_1)$ is drawn in this way. Comparing equations (20) and (9) shows that the shift of the household demand curve is ambiguous since the numerator is reduced by $-(1-\theta)\lambda\pi$ while the denominator is reduced by $-(\theta-c)\pi$. To emphasize the possibility of a lower equilibrium price even when the household demand price rises, the household demand curve is drawn with the demand price at the initial value of $s_{h1}=0$ greater than its previous level.

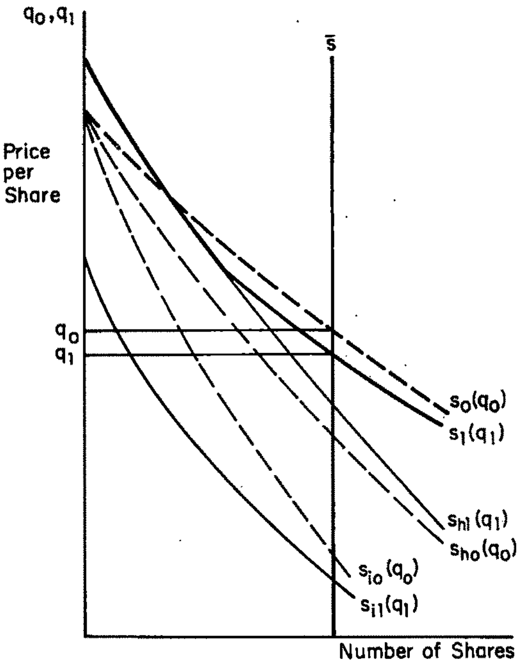


FIGURE 3

The new market demand curve $s_1(q_1)$ coincides with the household demand curve until a price is reached at which institutions are willing to hold some stock. Thereafter, the market demand curve is the sum of the two demands. The new equilibrium price occurs at a value of q_1 that is below the old equilibrium. Institutions reduce their shareholdings and individuals increase their shareholdings.

---

[20] See fn. 4 for evidence that this has been the historical experience in the United States.

Equations (20) to (22) can be used to calculate explicitly the values of q_1 and of the separate shareholdings. Combining (20) and (21) and using $\bar{s}-s_{h1}$ for s_{i1} yields an equation for the new shareholding by households:

$$(23)\quad s_{h1}=\frac{\delta_i\sigma_p(1-\theta)(1-\tau)\bar{s}(\theta-c)\pi}{\sigma_p(1-\theta)(1-\tau)[\delta_i+\delta_h]}$$

Consider first the new equilibrium when q_0 was equal to 1. Using the values of $\delta_h\sigma_p$ and $\delta_i\sigma_p$ implied by equations (16) and (17), and an inflation rate of $\pi=0.08$, equation (23) implies that $s_{h1}=754$, i.e., inflation causes households to increase their shareholdings from 700 shares to 754 shares.[21] Substituting this value into equation (20) and setting the historic-cost depreciation penalty at $\lambda=0.30$[22] yields $q_1=0.812$. The share price per unit of capital falls from one dollar to 81.2 cents. Note that the value of household shares is reduced by the inflation from \$700 billion to \$568 billion even though they hold an increased number of shares.

A lower initial value per share does not change the conclusion that the share price falls but does reduce the relative magnitude of the fall. More specifically, if $q_0=.8$,[23] equations (16), (17), and (23) imply that $s_{h1}=746$. Substituting into (20) yields $q_1=0.729$ or 91 percent of the initial price.

IV. Conclusion

The simple model developed in this paper conveys the idea of how a higher rate of inflation can cause a substantial reduction in the ratio of share prices to pretax earnings. The higher effective rate of tax on corporate income caused by historic-cost depreciation and the tax on the artificial capital gains caused by inflation both reduce the real net yield that investors receive per unit of capital. Although the real net yield on bonds is also reduced, for many shareowners this is outweighed by the fall in the equity yields.

The market equilibrium analysis examined the impact of inflation when both stocks and bonds are held by risk-averse investors in quite different tax situations. It also showed how the equilibrium ratio of share prices to earnings can fall even if the demand price per share for some individuals is actually increased by inflation.

Of course, the increase in the effective tax rate caused by inflation has not been the only adverse influence on the level of share prices during the last decade. The slowdown in productivity growth, the higher cost of energy, and the increased international competition have all reduced pretax profitability. Although there is no clear evidence of a permanent fall in profitability (see my paper with Summers, 1977), the transitory reduction may have caused some investors to project lower long-term pretax profitability. The higher tax rates on capital gains for high-income investors since 1969 further reduced after-tax profitability. An increase in uncertainty has also had an adverse effect on price-earnings ratios. One source of this greater uncertainty is the increasing ratio of debt to equity on corporate balance sheets. In addition, after a period of steady growth and low inflation, the events of the past decade have added uncertainty to any evaluation of the future. Finally, in considering the changes in the level of share prices over the past decade, it is important to recognize that the adverse effect of inflation has been perceived only slowly and imperfectly. Some investors have undoubtedly concluded incorrectly that even a steady rate of inflation would cause a continuing decline in the ratio of share prices to earnings. The share price level may therefore have overshot its equilibrium level.

A full understanding of the equilibrium relation between share prices and inflation requires extending the current analysis in a number of ways. The role of corporate debt and retained earnings should be included.[24]

---

[21] The other parameter values are $\tau=0.4$, $\theta=0.3$, $c=0.15$, and $\bar{s}=835$.

[22] See fn. 9.

[23] Auerbach and King show that under certain conditions the share price without inflation will be $q_0=(1-\theta)/(1-c)$ if the only shareholders are individuals with these tax rates. With our current tax values, this implies 0.82.

[24] An empirical analysis of corporate tax burdens with the existing corporate debt and retained earning

The possibility of individual investment in other assets like real estate should be recognized.[25] A more explicit portfolio model could derive asset demand equations from expected utility maximization and could recognize that some institutional holdings are really indirect ways for individuals to hold assets in a tax-favored way. Finally, the simplification that the capital stock remains constant should be replaced by a more dynamic model that recognizes the effect of inflation on capital accumulation.

---

shows that inflation raises the tax burden on equity investors as well as on total corporate sector capital; see my paper with Summers (1979).

[25] A model of the interaction of tax laws and inflation in determining the price of gold and land is presented in my forthcoming paper.

REFERENCES

A. Auerbach, "Share Valuation and Corporate Equity Policy," Nat. Bur. Econ. Res. work. paper no. 255, Cambridge, Mass. 1978.

D. Bradford, "The Incidence and Allocative Effects of a Tax on Corporate Distributions," Nat. Bur. Econ. Res. work. paper no. 349, Cambridge, Mass. 1979.

M. Feldstein, "The Effect of Inflation on the Prices of Land and Gold," *J. Publ. Econ.*, forthcoming.

______ **and O. Eckstein,** "The Fundamental Determinants of the Interest Rate," *Rev. Econ. Statist.*, Nov. 1970, *52*, 363–75.

______ **and L. Summers,** "Is the Rate of Profit Falling?," *Brookings Papers*, Washington 1977, *1*, 211–27.

______ **and** ______, "Inflation, Tax Rules and the Long-Term Interest Rate," *Brookings Papers*, Washington 1978, *1*, 61–99.

______ **and** ______, "Inflation and the Taxation of Capital Income in the Corporate Sector," *Nat. Tax J.*, Dec. 1979, *32*, 445–70.

______ **and** ______, "Inflation, Price-Earnings Ratios and the Cost of Capital," forthcoming.

______, **J. Green, and E. Sheshinski,** "Inflation and Taxes in a Growing Economy with Debt and Equity Finance," *J. Polit. Econ.*, Apr. 1978, *86*, Part II, S53–S70.

Irving Fisher, *The Theory of Interest*, New York 1930.

H. Hong, "Inflation and the Market Value of the Firm: Theory and Tests," *J. Finance*, Sept. 1977, *32*, 1031–48.

Mervyn King, *Public Policy and the Corporation*, London 1977.

J. Lintner, "Inflation and Common Stock Prices in a Cyclical Context," in *The New Realities of the Business Cycle*, Nat. Bur. Econ. Res. *53rd Annual Report*, New York 1973.

______, "Inflation and Security Returns," *J. Finance*, May 1975, *30*, 259–80.

F. Modigliani and R. Cohn, "Inflation, Rational Valuation, and the Market," *Finan. Analysts J.*, Mar. 1979, *35*, 3–23.

B. Motley, "Inflation and Common Stock Values: Comment," *J. Finance*, June 1969, *24*, 530–35.

C. Nelson, "Inflation and Rates of Return on Common Stocks," *J. Finance*, May 1976, *31*, 471–87.

J. Van Horne and W. Glassmire, Jr., "The Impact of Unanticipated Changes in Inflation on the Value of Common Stocks," *J. Finance*, Dec. 1972, *27*, 1081–92.

W. Yohe and D. Karnovsky, "Interest Rates and Price Level Changes, 1952–69," *Fed. Reserve Bank St. Louis*, Dec. 1969, *51*, 18–38.

A Theory of Just Regulation

By LI WAY LEE*

The regulation of prices, as most theories describe it, results in winners and losers.[1] Because someone is worse off as a result of regulation, the usual theories of regulation admit an arbitrary form of transitional inequity. In effect, they deny groups of individuals or firms the ability to organize or to find representation in the regulatory process to protect their interests. It is implausible to assume that issues of transitional gains and losses somehow fail to emerge in the regulation of prices, when they always loom large in other public policy areas such as taxation and public expenditure.[2]

In this paper, I present a theory of regulation that meets three requirements. First, the regulation makes both the consumers and the producers better off. Second, no cash transfer from the consumers to the producers takes place. Third, the regulator is "fair" (or just) in the determination of price.[3] When a regulation satisfies all three conditions, it is a just regulation.

I view regulation as basically an exchange process that involves the transfer of the monopoly-cartel enforcement function from the producer to the society. Furthermore, the transfer is reciprocated by the surrender of the power of price determination to a third party. Two conceptual steps are taken in creating this social institution. The first represents a social contract to transcend the market in order to bypass some of the loss of efficiency. The step involves an in-kind transfer (cartel enforcement) to the industry and a cash-equivalent surrender (pricing power) by the industry.[4] This exchange is opposed to the one involving cash transfer to the industry for an in-kind payment (increased output) to the public. The second step involves the arbitration of price.[5]

In this conceptual framework, then, a typical regulator acts as a social agent performing two functions: the enforcement of cartel order (or a privileged legal and economic status) and price arbitration. Regulatory agencies routinely perform entry control, auditing, etc., which are tasks that private cartels have to perform themselves in the absence of regulation. Price arbitration is a standard function of public utility regulators, as many structural and procedural characteristics of agencies reflect. For instance, Roger Noll observes that "regulatory agencies have been constructed to resemble high-level, collegial courts" and that "they appear to behave like these courts, emphasizing procedures and dealing with individual cases" (p. 7). The typical regulator's interest in the balance of interests, the making of compromises, or the facilitation of bargaining, is the same as that of any arbiter.[6]

---

*Wayne State University. I have benefited immensely from many hours of discussion with James Hamilton and from a detailed and insightful referee's report. My thanks for pointed comments also go to George Borts, Stephen Karlson, and Soc Bock Lee.

[1] See Richard Posner, and Bruce Owen and Ronald Braeutigam for brief but useful surveys.

[2] Gordon Tullock, Harold Hochman, and Owen and Braeutigam have considered the importance of transitional equity in policy making.

[3] I use just and fair interchangeably in this paper.

[4] For discussion of cash transfer in the context of regulation, see George Stigler (1971) pp. 4-5, and Gary Becker, pp. 247-48.

[5] Paul Joskow, p. 298, has stressed that the regulatory agency sets price, not rate of return.

[6] Posner argues: "Much regulation...may be the product of coalitions between the regulated industry and consumer groups..." (p. 351). Stigler envisions the theory of regulation as a theory of political equilibrium: "The final political equilibrium will be not that one party or the other 'wins' according to whose power is larger, but rather that an intermediate solution will be attained, where—at the margin—the strengths of the two parties are equal" (1975, pp. 138-39). Sam Peltzman has substantiated the proposition in his vote-maximization models of regulation. Also noteworthy is Claire Wilcox and William Shepherd's reference to regulation as "part of a larger social bargain" (p. 332) or as "part of a Basic Social Contract" (p. 348). In fact, regulatory agencies' effort at "balancing interest,"

The consensual notion of regulation is elaborated and applied below. The feasibility conditions for Pareto optimal regulations are explained in Section I. A model of the determination of a "just price" is presented in Section II. The model then becomes an integral part of the rest of the paper. In Section III, I discuss how the incentives for regulation are related to market structure. In Sections IV and V, the model is expanded to allow for interventions by antitrust agencies, consumer groups, and producer groups in the regulatory process.

I. The Efficiency Basis of Regulation

The essence of industrial power is control over supply that causes the market to generate less than the maximum amount of benefit to the society as a whole. Also, the control is an asset with a real cost of maintenance to the industry.

As a rule, it is difficult for producers in an industry to set up a collusive scheme, to police the cartel against price shading and related discounts, and, when antitrust laws are enforced vigorously, to evade detection and conviction. For the purposes of this paper, we need have only the broadest view of the cartel without regulation. In Figure 1, M is the cartel optimum without regulation.[7] The consumer's surplus at M is denoted by G^o and the profit at M is denoted by π^o.

The state has certain advantages over the private industry in the enforcement of private contracts, of which a cartel is a special case. The state has coercive powers such as imprisonment and the imposition of fines and taxation, which private industries could create only at greater costs (see Stigler, 1971, p. 4). Thus, the state may have economies of "size" in policing and auditing the conduct of a group.[8] The social mode of cartel enforcement may be much cheaper than the private mode.[9]

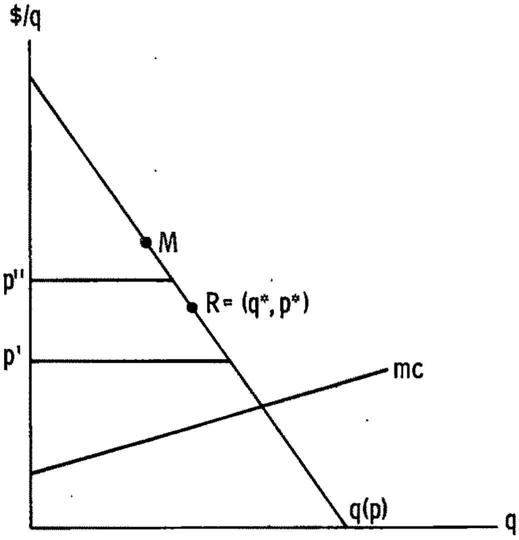


FIGURE 1

Regulation as the social mode of cartelization requires a public outlay (E). The outlay likely increases with the output level (q). It may depend on the market structure (h) as well. I assume that the cost is borne by consumers through general taxation.[10]

"making compromises" or "facilitating continual bargaining" is noted by almost all students of regulation. Only Peltzman and Gregg Jarrel, however, spell out exactly their visions of these social acrobatics.

[7] Uniform multipart pricing and discriminatory pricing will be briefly considered later.

[8] William Landes and Posner suggest that "...the economies of scale in [public law] enforcement appear to be concentrated in the investigative phase of enforcement (patrolling, auditing, etc.)..." (p. 30). They are not convinced, however, that public law enforcement in general has advantage over private law enforcement in terms of free-rider problems or economies of scale. Regarding entry deterrence, Harold Demsetz argues: "The techniques available to firms without the aid of government are not very specific to penalizing only entrants.... An investment by an industry to obtain government aid to monopolize is likely to yield much more control than the investment without the aid of government techniques" (1974, p. 181).

[9] I have assumed that the substitution of the enforcement function is complete. This may not be true. For example, the regulator may refuse to enforce product differentiation; consequently, the cartel still has to self-regulate quality rivalry. The extent of the substitution of the social for the private cartelization itself may be subject to bargaining. For any given aspect of cartel enforcement, the substitution may be less than total. Furthermore, the extent of substitution may vary from one aspect of the enforcement to another. I prefer the present approach for its simplicity.

[10] The relevant consideration should be that part of the social enforcement cost actually borne by con-

Then there exists a hypothetical *maximum* price at which the consumer's surplus is exactly the same as without regulation. This "social concession price" (p'' in Figure 1) must satisfy

$$(1) \qquad \int_{p''}^{\infty} q\,dp - E(q'';h) = G^o$$

where $q \equiv q(p)$ is the inverse demand function and $q'' \equiv q(p'')$. Similarly, now that the burden of enforcement has been transferred to the consumer, the cartel faces a hypothetical *minimum* price at which its profit is unchanged. This "industry concession price" (p' in Figure 1) must then satisfy

$$(2) \qquad p'q' - C(q') = \pi^o$$

where $q' \equiv q(p')$.

As illustrated in Figure 1, the two concession prices along with the demand curve define a potential, maximum efficiency gain from regulation.[11] In addition, p' and p'' establish the respective lower and upper bounds of the price to be set by the regulator.

For some price p^* between p' and p'', the consumer's surplus and the profit, are, respectively,

$$(3) \qquad G^* \equiv \int_{p^*}^{\infty} q(p)\,dp - E(q^*;h)$$

$$(4) \qquad \pi^* \equiv p^*q^* - C(q^*)$$

where $q^* \equiv q(p^*)$. It must also be true that both consumers and producers are better off:

$$(5) \quad G^* - G^o > 0 \qquad \text{and} \qquad \pi^* - \pi^o > 0$$

No cash transfer from the consumer to the producer need take place.

Note that the efficiency gain by regulation does *not* require a relative cost advantage of social enforcement. An efficiency gain still may result even if the social enforcement of the cartel has a cost disadvantage vis-à-vis the private enforcement, because regulation recovers part of the deadweight loss in the free market. Also note that the scheme explained above is sufficient but not necessary for Pareto optimal regulation without cash transfer. For example, regulation may involve the public enforcement of a multipart tariff or a discriminatory price structure that the industry enforced at high costs to itself (more on this in Section III).

In any case, it is part of the premise of my theory that industry profit will be higher and price lower than would be the case if the cartel were privately organized. Other theories *predict* differently. The public interest theory predicts that price and profit will be lower than if the cartel were administered privately (or enjoyed no preferential legal status). The capture theory predicts that both will be higher (see William Jordan). Peltzman's theory predicts that both will be lower (p. 223).

The four magnitudes, G^o, π^o, G^*, and π^* are the building blocks for the rest of the paper. I consider G^o and π^o to be functions of the structure of the industry (h), the antitrust enforcement (a), the consumer participation (x), and the industry participation (y). As to G^* and π^*, it already is clear that they depend on p^*. Further, G^* depends on h through the dependence of E on h. Later, I will assume that G^* and π^* are influenced by x and y. The nature of those dependencies will be explained as we go on. In sum, the four fundamental functions, with all their arguments written out, are: $G^o(h,a,x,y)$, $\pi^o(h,a,x,y)$, $G^*(p^*,h,x,y)$, and $\pi^*(p^*,x,y)$.

II. A Model of the Just Price

The socialization of cartel enforcement and the industry's relinquishment of pricing power pave the way for a price reset. Since price is crucial in determining the net gains from regulation, and inasmuch as pricing

---

sumers. However, this modification is not essential as long as the cost borne by consumers is proportional to the total social cost.

[11] For a well-behaved demand function, there are actually two prices that satisfy (1) and two other prices that satisfy (2). As long as the ranges defined by these two sets of prices overlap, there will be a gain from regulation.

power easily may be abused, the need for an impartial price setter is evident. In this section, I explore a model of just price.[12]

Observers of regulatory agencies often point out that regulators shun publicity and controversies, and that they try to be fair in their decisions. However, what would constitute a fair decision is not explicitly spelled out in statutory mandates or court interpretations. Regulatory agencies have no incentive to conform to any ideal principle of social justice. Rather, they try to fulfill the social expectation of what is fair. Thus, the agencies are subject to all the influences in the political economy when they arbitrate prices. As R. Duncan Luce and Howard Raiffa have observed in connection with arbitration: "We may suppose that the arbiter sincerely envisages his mission to be 'fairness'.... The arbiter can be assumed to want to suggest a solution which will seem 'reasonable,' both because he is sincere and because he may wish to be hired for such tasks in the future" (p. 121).[13] Arbitration usually reduces significantly the cost of bargaining. When it is binding, it also eliminates the possibility of failing to reach an agreement. For example, in a large-number context, such as in regulation, binding arbitration by a social agent overcomes the free-rider problem.

A remarkably simple and practicable concept of fairness is embodied in the Nash solution to bargaining (see Luce and Raiffa, pp. 124–37; Shubik, pp. 48–51). The solution satisfies four axioms of arbitration. We may state those axioms in our own terms. First, the regulated price must be invariant to linear transformation of the consumer's surplus and profit (invariant with respect to utility transformation). Second, there exist no other prices that make at least one group better off without making the other group worse off (Pareto optimality). Third, the regulator would not subsequently consider things that he did not concern himself with when he first determined the price (independence of irrelevant alternatives). Fourth, the factors that matter to the regulator are entirely economic (for example, cost structure and demand function); he is neither proconsumer nor proindustry (symmetry).

Nash proves that the utility function of the arbiter who adopts those axioms of arbitration is unique: it is the product of the net gains from bargaining. In my terms, the regulator who tries to be fair in the Nash sense above would determine a price p^* that maximizes

$$(G^* - G^o)(\pi^* - \pi^o) \qquad (6)$$

In this paper, I call that price the just price. I feel that the Nash axioms are believable in the context of regulation. At least, they do not seem totally arbitrary. As Luce and Raiffa have remarked: "[An arbiter] will try to satisfy some consistency requirement. In addition, as with most adjudicators, he will be anxious to defend his suggested solutions with some fairly good rationalization. All this means that he should be prepared to formulate and to defend the basic principles which lie behind his suggested compromises —they should not be completely arbitrary" (p. 121).[14] The Nash axioms give meaning to the observation that regulators are under pressure to make decisions that trigger the least amount of controversy.[15]

---

[12] Just price is an exceedingly appealing term, used frequently but seldom defined well. For example, J. de Graaff says: "I suggest that the only price a public enterprise or nationalized industry can be expected to set is what we may as well call a *just price—a price which is set with some regard for its effect on the distribution of wealth* as well as for its effect on the allocation of resources" (p. 155). Also, see David Winch.

[13] Regarding the regulator as an arbitrator helps account for many characteristics of the regulatory process. Consider, for example, the much observed and lamented fact that appointees to commissions are mostly politicians or lawyers (Noll, pp. 100, 108; Stigler, 1975, p. 163). By specialization, the expertise required of an arbitrator resides in these people rather than in economists. See Owen and Braeutigam for detailed discussion.

[14] See also Luce and Raiffa, pp. 128-34; Thomas Schelling, Appendix B; John Harsanyi (1956) p. 147, (1961) pp. 188-90; Amartya Sen, pp. 118-23; John Rawls, pp. 133-35. For applications of the Nash concept, see Harsanyi (1962) and Clive Bell and Pinhas Zusman.

[15] Noll has argued that one success indicator to a regulatory agency is "the failure of the courts or the Congress to override a decision by the regulators, on either procedural or substantive grounds" (p. 40). To

A notable feature of the Nash solution is its dependence on the so-called "threat positions" of the bargainers, namely, what they would receive respectively if the consensus broke down. In regulation, those positions simply are G^o and π^o, i.e., the consumer's surplus and profit without regulation. Thus, a Nash regulator is constantly aware of how things would be without regulation. (Whether actual regulators are is a question that can be answered only by checking one's own experience with predictions of my theory.) Last but not least, the absence of side payment in the solution reflects nicely my assumption that cash transfer is infeasible.

The surpluses and profits in this paper should be regarded as what the regulator perceives; they may be subject to various influences, as I will assume in later sections. Taking all extraneous influences as given, however, the regulator determines the price p^* that maximizes (6). The marginal condition for p^* is

$$(7)\quad (\pi^*-\pi^o)\frac{\partial G^*}{\partial p^*}+(G^*-G^o)\frac{\partial \pi^*}{\partial p^*}=0$$

From (3) and (4), it can be shown that

$$(8)\quad \frac{\partial G^*}{\partial p^*}=-q^*\left[1+\left(\frac{ME^*}{p^*}\right)\eta^*\right]$$

$$(9)\quad \frac{\partial \pi^*}{\partial p^*}=q^*\left[1+\left(\frac{p^*-MC^*}{p^*}\right)\eta^*\right]$$

where MC^* denotes the marginal cost of production (dC/dq), ME^* stands for the marginal cost of social enforcement $(\partial E/\partial q)$, and η^* is the demand elasticity, all evaluated at $R=(q^*,p^*)$ in Figure 1. The second-order condition is

$$(10)\quad D\equiv(G^*-G^o)\left(\frac{\partial^2\pi^*}{\partial p^{*2}}\right)+2\frac{\partial G^*}{\partial p^*}\frac{\partial \pi^*}{\partial p^*}+(\pi^*-\pi^o)\left(\frac{\partial^2 G^*}{\partial p^{*2}}\right)<0$$

From (7) (or the Pareto optimality of the Nash solution), it follows that $\partial G^*/\partial p^*$ and $\partial\pi^*/\partial p^*$ take on opposite signs or that they both vanish. In the latter case, the regulated price maximizes both the consumer's surplus and profit. When, as is more likely, G^* and π^* vary with p^* in opposite directions, one of the following two Pareto optimality conditions must hold

$$(11)\quad \partial G^*/\partial p^*<0 \quad\text{and}\quad \partial\pi^*/\partial p^*>0$$

or

$$(12)\quad \partial G^*/\partial p^*>0 \quad\text{and}\quad \partial\pi^*/\partial p^*<0$$

Pareto optimality condition (11) is the more familiar of the two. It indicates that, at the regulatory optimum, the consumer would benefit from a higher output at a lower price, while the producer would benefit from a lower output at a higher price. This condition must follow when E is fixed (i.e., $ME=0$). Pareto optimality condition (12) is unusual. It means that the consumer would benefit from a lower output at a higher price, while the producer would benefit from a higher output at a lower price. This condition applies when the maximum consumer's

---

this he adds a second success indicator: "continued operation of the regulated sector" (p. 40). My assumption is also similar to Joskow's in his theory of the regulatory agency which involves two "primary actors": the firm and the regulatory agency. Consumer interest is one of the "potential actors." Rate and return structures are determined on certain norms of "fairness." These norms are adopted so as to "minimize conflict and criticism appearing as 'signals' from the economic and social environment in which they operate, subject to binding legal and procedural constraints imposed by the legislature and the courts" (p. 297). His theory assigns a very passive role to the consumer group. Consumers do not intervene if price remains constant or drops; rate of return is of no interest to them. Owen and Braeutigam lay great emphasis on the fairness in regulation. They argue that the reliance of the regulatory process on administrative rules makes possible effective representation of interest groups, and that the power to delay wielded by interest groups in regulation tends "to slow down the rate at which the free market redistributes income, thus reducing the risks faced by voters" (p. 26). Regulation guarantees fair treatment because "[t]he essence of administrative law is that the decision-making process must conform to reasonable standard of fairness..." (p. 22).

surplus occurs at a lower output level than the profit optimum, with R lying between the maxima. It is a logical possibility following from the social enforcement of cartel (see (3)). In this paper, I assume that condition (11) holds. Condition (12) is ruled out as *empirically* implausible in the regulation of price.[16]

The marginal condition (7) may be written explicitly as

$$(13)\qquad S^* = \left[1+\left(\frac{p^*-MC^*}{p^*}\right)\eta^*\right] \Big/ \left[1+\left(\frac{ME^*}{p^*}\right)\eta^*\right]$$

where $S^*(\equiv(\pi^*-\pi^o)/(G^*-G^o))$ stands for the net-gains ratio. By the assumption of condition (11), both the numerator and denominator are positive. One can see that, at the regulatory optimum, the net-gains ratio is systematically related to the marginal social surplus of output (MSS^*):

$$(14)\qquad MSS^* \equiv p^* - (MC^* + ME^*)$$

Consider the following possibilities:

(a) When the marginal social surplus is negative ($MSS^* < 0$), producers gain more from regulation than consumers ($S^* > 1$).

(b) When the marginal social surplus is zero ($MSS^* = 0$), the net gains are equal ($S^* = 1$). Such is the case, for example, when regulation maximizes simultaneously the consumer's surplus and profit.[17]

(c) When the marginal social surplus is positive ($MSS^* > 0$), consumers gain more from regulation than producers, ($S^* < 1$).

Therefore, regulation will likely entail deadweight loss to the extent that the marginal social surplus at the regulatory optimum is nonzero. Zero deadweight loss results only when consumer's surplus and profit are maximized simultaneously at the regulatory optimum.

III. Market Structure and the Incentive for Regulation

If a consumer were to pick out an industry for regulation from a continuum of industries that differ only in their structure, having identical demand and cost conditions, which one would it be?[18] Which industry has the strongest incentive for regulation? Answers to these questions would provide us with some clue as to which industries are likely to be regulated.[19]

Assume that one can define market structure (h) by the concentration in the market. Choose the concentration measure that relates most closely to the degree of collusion in the industry.[20] The continuum of industries then is indexed by h. Associated with each h are a pair, G^o and π^o. Following the concentration-collusion doctrine, assume further that, across the continuum, industry profit without regulation is positively related to industrial concentration, and that the consumer's surplus without regulation is negatively related to the industrial concentration.[21] It also seems reasonable to assume that the cost of social enforcement of a cartel decreases with concentration. For, the smaller is the number of firms, the lower would be the cost of controlling their behavior. In symbols, these assumptions are

---

[16] In other areas of public policy, however, condition (12) is not implausible. When overburdened by taxes, consumers often wish to have lower governmental output at a higher tax price.

[17] The simultaneous maxima occur only if $(ME^*/p^*)=(p^*-MC^*)/p^*=-1/\eta^*$ at R, as can be seen from (8) and (9).

[18] I pose the question in this peculiar way to preclude the effect of size on incentives: across industries of comparable concentration, those with large sale volume will be more likely regulated than small industries, *ceteris paribus*. It also suggests that as an industry experiences secular growth in demand, the attraction of regulation will become greater.

[19] Posner, for example, has remarked that "...the economic theory has not been refined to the point where it enables us to predict specific industries in which regulation will be found. That is because the theory does not tell us what (under various conditions) is the number of members of a coalition that maximizes the likelihood of regulation" (p. 347).

[20] See Stigler (1964), and Robert Dansby and Robert Willig on the rationales for concentration indices.

[21] Demsetz discusses the alternative hypothesis that concentration is related to lower cost of production. When this is true, it is possible that consumer's surplus without regulation is positively related to industrial concentration.

$$(15)\quad \frac{\partial \pi^o}{\partial h}>0;\ \frac{\partial G^o}{\partial h}<0;\ \frac{\partial G^*}{\partial h}\left(=-\frac{\partial E}{\partial h}\right)>0$$

Market structure is assumed to have no direct effect of π^*, the profit of the industry with regulation.

Suppose further that if the continuum of industry were regulated, they all would be Pareto optimal. Then there exists a continuum of regulated price. How does the regulated price vary with structure, *ceteris paribus*? By totally differentiating (7) with respect to h, we obtain

$$(16)\quad \frac{\partial p^*}{\partial h}=-\left\{\frac{\partial \pi^*}{\partial p^*}\left(\frac{\partial G^*}{\partial h}-\frac{\partial G^o}{\partial h}\right)-\frac{\partial G^*}{\partial p^*}\frac{\partial \pi^o}{\partial h}+(\pi^*-\pi^o)\frac{\partial^2 G^*}{\partial p^*\partial h}\right\}D^{-1}$$

Because it follows from the concentration-collusion doctrine that high concentration endows the industry with a strong threat position relative to that of the consumer, and because high concentration, for a given price, benefits consumers by reducing the cost of enforcement (see (15)), the first two terms tend to make the regulated price rise with concentration. The third term may work in the opposite direction because high concentration intensifies the depressing effect of price on consumer's surplus, as the saving in the social enforcement cost that accompanies the reduction in output due to higher price is low for a highly concentrated market.[22] If, as I feel is likely, market structure affects only the fixed-cost component of E, the third term would be zero. In this case, the sign of (16) is unambiguously positive, implying that highly concentrated industries, *ceteris paribus*, have high regulated prices. If, on the other hand, market concentration reduces E by shifting ME downward, the sign of (16) would be indeterminate. In the latter case, concentration may exert an overall negative influence on regulated price for certain ranges of market structure. I suspect that this is unlikely. But then the sign is best investigated empirically.

Turning now to the incentives, I associate the net gains from regulation, (G^*-G^o) and $(\pi^*-\pi^o)$, with the consumer's incentive and the producer's incentive for regulation, respectively. I first examine the dependence of the consumer's incentive for regulation on market structure, after which I examine that of the industry's.

The consumer's incentive for regulation varies with market structure in the following way:

$$(17)\quad \frac{\partial(G^*-G^o)}{\partial h}=\frac{\partial G^*}{\partial p^*}\frac{\partial p^*}{\partial h}-\frac{\partial E}{\partial h}-\frac{\partial G^o}{\partial h}$$

The various terms in the expression are defined by (8), (16), and (15). Figure 2 depicts the simplest behavior of the consumer's incentive as a function of concentration. Note that the incentive reaches a maximum, defined by $\partial(G^*-G^o)/\partial h=0$. Thus, one may in principle determine the concentration most conducive to regulation from the consumer's point of view.

The industry's incentive for regulation is also positively related to the net gain from regulation. It varies with concentration according to

$$(18)\quad \frac{\partial(\pi^*-\pi^o)}{\partial h}=\frac{\partial \pi^*}{\partial p^*}\frac{\partial p^*}{\partial h}-\frac{\partial \pi^o}{\partial h}$$

The various terms in the expressions are defined by (9), (16), and (15). Again, Figure 2 describes the simplest behavior of (18). The maximum industry incentive for regulation is reached at a concentration defined by $\partial(\pi^*-\pi^o)/\partial h=0$. Note that the optimal concentration for the consumer in general differs from that for the industry.

It might seem unlikely that the optimal concentrations are extremely high (monopolistic), or low (atomistic). For both monopolistic and atomistic markets, no significant cost of maintaining industrial power is in-

---

[22] My reasoning follows from differentiating (8): $\partial(\partial G^*/\partial p^*)/\partial h=-(\partial ME^*/\partial h)(\partial q^*/\partial p^*)<0$.

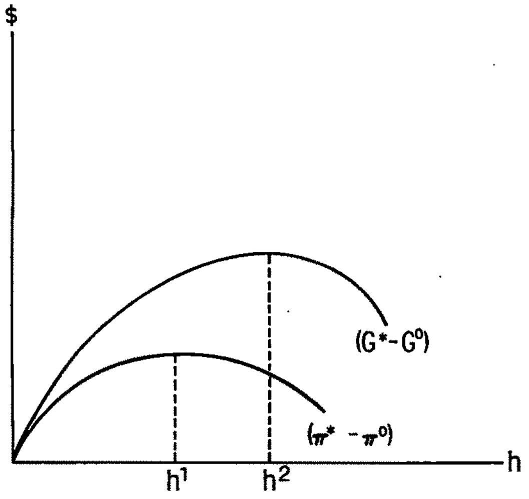


FIGURE 2

curred by producers. Hence, Pareto optimal regulation would seem improbable. However, I contend below that the bargaining theory of regulation is capable of explaining regulations of monopolistic and atomistic markets when we relax certain assumptions of the theory thus far.

I have maintained, up to this point, that Pareto optimal regulation can be based on the socialization of the cost of maintaining industrial power and that the industry practices simple pricing, with or without regulation. These assumptions are not necessary for Pareto optimal regulation.

A monopolist in the truest sense of the word may submit to regulation if regulation can cut down the *cost of production*. For example, if the cost of production could be reduced through the use of eminent domain, then a monopolistic producer might agree to regulation. My theory still applies even though here there is no pecuniary burden of enforcement on the consumer. Alternatively, the regulation of a natural monopoly may involve the offer to the monopolist of a schedule of discriminatory prices or a multipart tariff entirely below the monopolist's, in return for the control over price. Without regulation, the natural monopolist either finds price discrimination unprofitable to practice (due to cost of administration, resale, or antitrust), or practices it at high costs. Regulation benefits the natural monopolist not only because his cost of output goes down by the amount of the cost of enforcing a price schedule, but also because the revenue possibly goes up as the marginal price is brought more in line with marginal cost without fear of losing revenue on intramarginal units. The regulation of natural monopoly may benefit consumers because they individually pay lower (though perhaps still multipart or discriminatory) prices and now consume more, and because new consumers can be brought into the market due to the lower, regulated price schedule.

Similarly, one need not always fall back to some political theory of regulation to explain the regulation of atomistic industries (for example, trucking). The regulation of some atomistic markets can be regarded as a consequence of the social consensus reached between a cartel and consumers at an earlier time. To the extent that the social contract involves a transfer of the cartelization function to the public, the restriction of entry will be regarded as an integral part of the contract. The effect of the appearance of a good substitute on the demand as perceived by the cartel is exactly the same in nature, if not in degree, as that of the entry of a producer of an identical product. The cartel naturally would bargain for the restriction of the competition from close substitutes. Further, the cartel would demand that regulation be imposed on a substitute industry in its incipiency, namely, before its full competitiveness could be developed. The regulation of close substitutes would very early maintain sufficiently high prices for the substitutes, or attempt to hinder nonprice competition by the substitutes. It is no wonder that regulation tends to spread from an established market over to developing, related ones. One can also understand why established industry prefers that the regulatory power over related markets be centralized in a single regulatory agency. Hence, when the railroad cartel and the consumers submitted to regulation, they had committed themselves to the regulation of close substitutes if and when they appear. When trucking had proven itself to be an effective competitor to railroads, the regulator responded by fulfilling an implicit contractual obligation to the railroad cartel.

In sum, while my analysis makes the most attractive market structure for Pareto optimal regulation depend on the details of regulation itself, it does imply that the most *unattractive* one, from both the consumer's and the producer's viewpoints, is an atomistic market that poses no threat to any established, regulated market. These implications differ from those of other theories. The public interest theory predicts that industries with natural monopolistic structure will be most attractive to regulate. The capture theory has no definite prediction of industries most likely to obtain regulation, presumably because of the important roles assigned to political variables. Peltzman, in his generalization on the capture theory, finds that "[e]ither natural monopolistic or natural competitive industries are more *politically* attractive to regulate than an oligopolistic hybrid" (pp. 223–24, emphasis added). My analysis, I should repeat, deals with the *economic* incentives for regulation.

IV. The Interaction of Antitrust and Regulation

In many markets, regulatory statutes make special provisions to exempt specific or general actions of the regulated industries from antitrust prosecution as long as the actions have the approval of the regulators (see Walter Adams). However, antitrust agencies still may participate in regulatory proceedings as intervenors and may litigate (see Phillip Areeda). How does antitrust policy affect the regulatory outcome?

By attacking industrial collusion in general, antitrust affects the regulatory outcome through two channels in my model.[23] 1) Antitrust decreases the profit of the industry without regulation, thereby weakening the industry's bargaining position in regulation. 2) It increases the consumer's surplus without regulation, thereby strengthening the bargaining position of the consumer.[24] Denoting the antitrust activity level by a composite index of inputs, a, we write these assumptions in symbols as

$$(19)\qquad \frac{\partial \pi^o}{\partial a} < 0; \qquad \frac{\partial G^o}{\partial a} > 0$$

To see how these affect the regulated price, differentiate (10) totally with respect to a and obtain

$$(20)\qquad \frac{\partial p^*}{\partial a} = \left\{ \frac{\partial G^*}{\partial p^*}\frac{\partial \pi^o}{\partial a} + \frac{\partial \pi^*}{\partial p^*}\frac{\partial G^o}{\partial a} \right\} D^{-1} < 0$$

The two channels of influence are represented by the two terms in the numerator, respectively. Since the sign is negative (see (8), (9), and (19)), I conclude that stepped-up antitrust activity will lower prices. The result is consistent with the impression that the application of antitrust laws in a regulated market can "stimulate substantial adjustments by a regulatory agency" (Edward Zimmerman, p. 382).[25]

The role of antitrust in regulation as described above is entirely indirect. The antitrust agency need not carry out a program directed at the regulated industry other than perhaps participation in the regulatory process as an intervenor. Inherent in the Nash solution is the implication that the regulator keeps himself informed of the general level of antitrust activity in the economy and of the current gist of antitrust philosophy. Therefore, the regulator has an idea of the threat positions, π^o and G^o, as they are related to a. (Note that the threat positions would realize instantaneously once the regulatory consensus breaks down; one need not traverse backward in time to find those positions.) Antitrust activity undirected at the regulated industry still would depress the regulated price.

---

[23] In my earlier paper, I treat some models of interaction of cartel and antitrust.

[24] In addition to those two effects, antitrust intervention also may affect directly the cost of social enforcement. Thus, E may be written as $E(q, a)$. I feel that such consideration is secondary and have chosen to ignore it.

[25] G. William Schwert, pp. 143-44, has found evidence that the pressure put by the Antitrust Division on the Securities and Exchange Commission to abandon price fixing among stockbrokers brought about substantial reductions in New York Stock Exchange seat prices.

The analysis above has broader application. I have said in Section III that the appearance of substitutes (or events leading to lower barrier to entry) would create demand on the part of the regulated industry for their regulation. The analysis above implies, however, that the regulation of substitutes never can be perfectly effective in protecting the interest of the established industry. The threat position of the industry weakens as it becomes evident to all that substitutes are available and could compete effectively if regulations were not in force. It follows that the regulator would, in fairness, determine a lower price for the established industry. Hence, petitions for entry, even if consistently denied, and the appearance of close substitutes, even if immediately regulated, still would lower the regulated price of the established regulated industry.

How does antitrust policy affect the incentives for regulation in a regulated industry? I examine the effect on the incentive for the consumer group first. The consumer's net gain from regulation (G^*-G^o) is a function of the antitrust activity, a. The marginal effect of a on consumer incentive is

$$(21)\qquad \frac{\partial(G^*-G^o)}{\partial a}=\frac{\partial G^*}{\partial p^*}\frac{\partial p^*}{\partial a}-\frac{\partial G^o}{\partial a}$$

From (8), (20), and (19), it is seen that both terms on the right-hand side of the equation are positive. This means that the consumer surpluses, G^* and G^o, always increase with antitrust. However, the difference between the two, which is the consumer's incentive for regulation, may first increase with antitrust activity and then decrease with it. The optimal antitrust activity on the regulated market from the consumer's point of view is defined by $\partial(G^*-G^o)/\partial a=0$.

Next, note that the producer gain from regulation $(\pi^*-\pi^o)$ is also a function of a:

$$(22)\qquad \frac{\partial(\pi^*-\pi^o)}{\partial a}=\frac{\partial \pi^*}{\partial p^*}\frac{\partial p^*}{\partial a}-\frac{\partial \pi^o}{\partial a}$$

From (9), (20), and (19), it is seen that both terms on the right-hand side of the equation are negative. What this means is that, although π^* and π^o decrease with antitrust, their difference, which is the industry's incentive for regulation, may initially increase with antitrust activity and then decrease.

These observations point to the possible existence of a range of antitrust activity within which both the consumer and the producer will have a growing desire for existing regulation. As the antitrust activity increases beyond this range, there may be conflict in that one party's desire grows while the other's diminishes. Finally, as the antitrust activity increases further, the desire for regulation diminishes for both. These ranges can be illustrated by Figure 2, where h is replaced by a, h^1 by a_1, and so forth. For $a<a_1$, both incentives grow with a. For $a_1<a<a_2$, the incentives vary in opposite directions. For $a>a_2$, both incentives diminish with a.

Finally, it may also be seen from (21) and (22) that, as antitrust activity grows, it would be in the interest of the industry and to the detriment of the consumers *somehow* to exempt totally the regulated industry from antitrust prosecution (i.e., to make $\partial p^*/\partial a=0$). In this way, the interest of the industry in regulation would grow continuously with antitrust, since without the shield of regulation, the industry would have been in great trouble.

The analysis of market structure and incentives for regulation in Section III also implies that antitrust policy may affect the regulatory outcome through a program aimed at changing the structure of the regulated industry. For instance, deconcentration of the regulated industry, with no side effect on costs and demand conditions, may (but also may not) decrease the regulated price in accordance to (16).

The calculation of the effect of antitrust enforcement on the regulated price $(\partial p^*/\partial a)$ shows the redistributive impact of antitrust; it provides no information about the relationship between antitrust and efficiency. To discuss the latter, one needs a well-specified model of regulation-antitrust interaction. In the model examined below, I assume that the antitrust agency takes the reaction function of the regulated price to antitrust activity as given and determines the level of antitrust activity to maximize its own objective. I further assume that the antitrust agency maximizes the net social surplus

$$U^*=G^*+\pi^*-A(a) \tag{23}$$

where $A(a)$ is the opportunity cost of a. The "efficiency" bias introduced here is obvious. The regulator is assumed to be concerned about the fairness of the division of social surplus, while the antitrust agency is assumed to be concerned about the size of the social surplus. This behavioral assumption admittedly has a weaker empirical foundation than that about the regulator. Antitrust authorities may be concerned with "fair competition" in some sense, instead of with efficiency *per se*. On the other hand, existing empirical evidence of the antitrust agency's behavior, when properly interpreted, do not show gross mismanagement of antitrust resources (see my earlier paper).

The first-order condition for the optimal antitrust policy is obtained by maximizing U^* with respect to the antitrust activity:

$$[p^*-(MC^*+ME^*)]\frac{\partial q^*}{\partial p^*}\frac{\partial p^*}{\partial a}=\frac{\partial A}{\partial a} \tag{24}$$

where the interaction term $\partial p^*/\partial a$ is defined by (20). We may express the condition in terms of elasticities:

$$\frac{MSS^*}{p^*}=\frac{A}{p^*q^*}\,\frac{\theta^*}{\phi^*\eta^*} \tag{25}$$

where θ^* is the elasticity of A with respect to a, $\phi^*(<0)$ is the elasticity of the regulated price p^* with respect to antitrust spending a, and $\eta^*(<0)$ is the demand elasticity, all evaluated at (p^*, q^*).

Conditions (7) and (24) form a model of antitrust and regulation interaction, in which the antitrust agency acts as the leader and the regulatory agency as the follower. Several implications of the antitrust regulation condition (25) are noteworthy. First, the marginal social surplus remains positive. It follows that consumers gain more from regulation than producers (recall the discussion below (14)). The antitrust agency intervening in a regulated industry acts *as if* it were proconsumer. Second, the condition above implies that the more recalcitrant is the regulated price to antitrust (i.e., lower ϕ^*), *ceteris paribus*, the higher is the marginal social surplus at the antitrust regulation equilibrium. Third, the more inelastic is the demand of an industry, *ceteris paribus*, the higher is the marginal social surplus at the equilibrium.

V. Information and Intervenors

In the regulatory process described thus far, consumers and producers have had no active role to play. In reality, however, consumers and producers often participate in regulatory proceedings to influence the price. Further, their presence seems to matter.[26]

There are two roles for intervenors in the regulatory processes. One role is political: to press for "preferential" treatment. The other is economic: to provide the regulator with information. I do not consider the political role of intervenors for two reasons, the first being that the political approach in effect requires that the Nash solution to arbitration be abandoned.[27] The second and more

---

[26] Joskow (1972) has found in a sample of electricity and gas company rate cases in New York that the presence of cost of capital testimony by the companies raised the allowed rate of return by 0.40 percentage points on the average. On the other hand, the presence of cost of capital testimony by consumer-interest intervenors reduced the allowed rate of return by 0.172 percentage points on the average.

[27] One way of incorporating the political role of intervenors is to remove the assumption of symmetry in the Nash solution concept of arbitration. Let x and y, composite indices of inputs, be the participation activity levels of the consumers' group and the producers' group, respectively. The regulator's objective may be hypothesized to be the determination of a price that maximizes the following function

$$(G^*-G^o)^{f(x)}(\pi^*-\pi^o)^{g(y)}$$

where $f(x)$ and $g(y)$ are increasing functions of x and y. The price that maximizes the above function satisfies all the Nash assumptions but symmetry. The symmetry assumption is satisfied only if $f(x)$ and $g(y)$ turn out to be equal.

important reason is that I have found that the implications of the political approach do not qualitatively differ from those of the economic approach.

Most regulators operate with imperfect information. In particular, a regulator may lack precise knowledge of the factors underlying the consumer's surplus and profit, with or without regulation. To the extent that a regulator makes decisions on the basis of what he knows, intervenors may affect the regulatory outcome by providing information to the regulator *strategically*. (Meanwhile, I assume that the regulator still lives by his Nash principles of fairness.) Let x and y, composite indices of inputs, be the participation activities of the consumer and the producer groups, respectively. The effects of x and y on p^* are given by the following derivatives obtained from (7) by total differentiations. For a consumer intervenor,

$$(26)\quad \frac{\partial p^*}{\partial x}=-\left\{\frac{\partial \pi^*}{\partial p^*}\left(\frac{\partial G^*}{\partial x}-\frac{\partial G^o}{\partial x}\right)+\frac{\partial G^*}{\partial p^*}\left(\frac{\partial \pi^*}{\partial x}-\frac{\partial \pi^o}{\partial x}\right)+(\pi^*-\pi^o)\frac{\partial^2 G^*}{\partial p^*\partial x}+(G^*-G^o)\frac{\partial^2 \pi^*}{\partial p^*\partial x}\right\}D^{-1}$$

The counterpart derivative for an industry intervenor is the expression above with y replacing x.[28] One may predict the signs of these expressions by anticipating how the intervenors typically would argue in a rate hearing before the regulator.

When a consumer intervenor testifies, the following is the likely substance of his testimony:

1) He advises the regulator of the "inadequate" benefit that consumers receive as a result of regulation: $\partial(G^*-G^o)/\partial x<0$.

2) He argues that the current profit of the producer is "excessive": $\partial(\pi^*-\pi^o)/\partial x>0$.

3) He emphasizes that at present, lowering the price would bring "significant" benefit to consumers: $\partial(\partial G^*/\partial p^*)/\partial x<0$.

4) He points out that the effect of lowering the price on the producer's profit would be "insignificant": $\partial(\partial\pi^*/\partial p^*)/\partial x<0$.

Turning to (26), we see that those points of advocacy precisely would make the sign of $\partial p^*/\partial x$ negative. Thus, the advocacy by intervenors on behalf of consumers would lower the regulated price. The presence of the producer representatives arguing just the opposite would raise the regulated price ($\partial p^*/\partial y>0$).

To formulate the intervenors' behavior, we need to know their objectives and how they each expect other actors in regulation to react (i.e., what they do and do not take as given). The following formulation is simple and intuitive. The consumer group intervenes in the regulatory process to maximize the consumer's surplus (G^*); the producer group intervenes to maximize profit (π^*). The consumer intervenors observe only the reaction of price to their own participation activity level, taking as given the producers' and the antitrust agency's activities. As to the producer intervenors, they also take into account the reaction of price to their own participation activity, expecting the consumer intervenors and the antitrust agency's activities to remain fixed. Further, it is only reasonable to assume that the consumer intervenors themselves do not see x as an argument of G^* and that the producer intervenors themselves do not see y as an argument of π^*. Those arguments were introduced to represent the effects of persuasion on regulatory outcome. Intervenors need not persuade themselves.

Now the conditions for optimal participation by the consumer group and the producer group can be derived. For the consumer group, it maximizes $(G^*-X(x))$ subject to (13) by a suitable choice of x, where X stands for the opportunity cost of x. The marginal condition for optimal consumer intervention is

$$(27)\quad \frac{\partial G^*}{\partial p^*}\frac{\partial p^*}{\partial x}=\frac{dX}{dx}$$

---

[28] In deriving this expression, I ignore possible effects of x (and y) on E.

The producer group maximizes $(\pi^* - Y(y))$ subject to (13) by a suitable choice of y, where Y is the opportunity cost of y. The marginal condition for optimal producer intervention is

$$\frac{\partial \pi^*}{\partial p^*}\frac{\partial p^*}{\partial y} = \frac{dY}{dy} \tag{28}$$

For the second-order conditions, diminishing marginal returns to persuasion $(\partial^2 p^*/\partial x^2, \partial^2 p^*/\partial y^2 < 0)$ and increasing marginal costs of persuasion $(\partial^2 X/\partial x^2, \partial^2 Y/\partial y^2 > 0)$ would be sufficient. In the marginal conditions, $\partial p^*/\partial x$ and $\partial p^*/\partial y$ are reaction terms defined by (26) and the counterpart for y.

All in all, there are four actors in regulation: the regulator, the antitrust agency, the consumer group, and the industry group. Each behaves in a well-specified way. The regulator behaves in the Cournot manner. The remaining three behave toward the regulator in the von Stackelberg manner, and toward each other in the Cournot manner. The interaction of these actors is represented by a system of four equations: (7), (24), (27), and (28). The solution to this system of equations constitutes the regulatory equilibrium.

Instead of reporting on the full comparative static properties of the regulatory equilibrium, here I remark on some simple but useful properties of the reaction functions. One example should suffice. Equation (27) may be regarded as the reaction function of the consumer intervenors with respect to the producer intervenors (and the antitrust agency). We ask: How do the consumer intervenors behave as the producer intervenors increase their activity? Take the total differential equation of (27) with respect to x and y (assuming $da = 0$) and rearrange:

$$\left[\frac{\partial G^*}{\partial p^*}\frac{\partial^2 p^*}{\partial x^2} + \frac{\partial^2 G^*}{\partial p^* \partial x} - \frac{d^2 X}{dx^2}\right] dx = -\left[\frac{\partial G^*}{\partial p^*}\frac{\partial^2 p^*}{\partial x \partial y} + \frac{\partial}{\partial y}\left(\frac{\partial G^*}{\partial p^*}\right)\right] dy \tag{29}$$

The bracketed term on the left-hand side is the second-order condition of (27); therefore it is negative. The one on the right-hand side consists of two mathematically complicated terms, on which, however, one may form reasonable expectations. We may expect that more producer-intervenors' activity, other things being equal, diminishes the marginal effectiveness of consumer intervenors $(\partial(\partial p^*/\partial x)/\partial y < 0)$, and aggravates the adverse effect of price on the consumer's surplus $(\partial(\partial G^*/\partial p^*)/\partial y > 0$.[29] It then follows that $\partial x/\partial y$ is positive, implying that as the producer-intervening activity increases, the consumer-intervening activity also increases. A similar analysis of equation (28) leads us to the reasonable conclusion that the producer-interest intervenors also react to increasing consumer-intervening activity by increasing their own activity.

Thus, we expect that measures that exclusively reduce the marginal cost of consumer intervention in regulatory proceedings, such as those facilitating the process of aggregating individual consumer interests into a collective interest, will lead to heightened consumer participation as well as producer participation, even though the regulated price may be lowered. Further, measures that reduce the cost of consumer as well as producer participation simply would increase the intervening activities of both, while possibly leaving the price unchanged.

VI. Conclusion

In this paper, I have discussed regulation as a collective choice to transcend the market in order to realize efficiency gains. This is accomplished through the socialization of cartel enforcement and the institution of social arbitration in price determination. The

---

[29] The persuasion variable y affects the marginal rate of change in G^* with respect to p^* in two ways:

$$\partial\left(\frac{\partial G^*}{\partial p^*}\right)\Big/\partial y = \frac{\partial^2 G^*}{\partial p^{*2}}\frac{\partial p^*}{\partial y} + \frac{\partial^2 G^*}{\partial p^* \partial y}$$

First, y raises p^*. Second, y increases $\partial G^*/\partial p^*$ by instilling the notion that price does not severely affect consumer's surplus (that is, the second term on the right-hand side is positive).

usual assumption that some interest group necessarily loses or wins in regulation is avoided. Both the consumer group and the producer group gain in my model. Moreover, the arrangement does not rely on open cash transfer from the consumers to the producer, which has been politically unpopular. The arrangement involves a transfer in kind from the public to the producer (cartel enforcement) and a cash transfer from the producer to the public (lower price). Thus, the arrangement not only is mutually beneficial to the consumer and the producer, but also possesses a high degree of political feasibility.

The conflict that does arise in my model has to do with how the gain from regulation is to be shared. Fairness becomes the overriding concern in the process.

I believe that the Nash solution to arbitration captures well the social reality that what is fair depends on the threat advantages of the groups in conflict. The deliberate vagueness in statutory mandates regarding the meaning of just price or "fair return," and the notorious inability of courts to define exactly what they are, amount to a tacit admission that the very concepts are undefinable in general. They are to be defined within specific contexts. What the Nash solution implies is that the groups themselves will ensure that the outcome reasonably reflects their relative threat advantages in a specific context. Therefore, the pressure tactics being applied to regulatory agencies may be simply informative.

REFERENCES

W. Adams, "Exemptions from Antitrust: Their Extent and Rationale," in Almarin Phillips, ed., *Perspectives on Antitrust Policy,* Princeton 1965.

P. E. Areeda, "Antitrust Laws and Public Utility Regulation," *Bell J. Econ.*, Spring 1972, *3*, 42–57.

G. Becker, "Comment (on Peltzman)," *J. Law Econ.*, Aug. 1976, *19*, 245–48.

C. Bell and P. Zusman, "A Bargaining Theoretical Approach to Cropsharing Contracts," *Amer. Econ. Rev.*, Sept. 1976, *66*, 578–88.

J. M. Buchanan and G. Tullock, "The 'Dead Hand' of Monopoly," *Antitrust Law Econ. Rev.*, Summer 1968, *1*, 85–96.

R. E. Dansby and R. D. Willig, "Industry Performance Gradient Indexes," *Amer. Econ. Rev.*, June 1979, *69*, 249–60.

H. Demsetz, "Industry Structure, Market Rivalry, and Public Policy," *J. Law Econ.*, Apr. 1973, *16*, 1–9.

______, "Two Systems of Belief about Monopoly," in Harvey J. Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston 1974.

J. de V. Graaff, *Theoretical Welfare Economics*, London 1967.

J. C. Harsanyi, "Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories," *Econometrica*, Apr. 1956, *24*, 144–57.

______, "On the Rationality Postulates Underlying the Theory of Cooperative Games," *J. Conflict Resolution*, No. 2, 1961, *5*, 179–96.

______, "Measurement of Social Power, Opportunity Costs and the Theory of Two-Person Bargaining Games," *Behav. Sci.*, Jan. 1962, *7*, 67–80.

H. M. Hochman, "Rule Change and Transitional Equity," in his and George E. Peterson, eds., *Redistribution Through Public Choice*, New York 1974, 320–41.

G. A. Jarrell, "The Demand for State Regulation of the Electric Utility Industry," *J. Law Econ.*, Oct. 1978, *21*, 269–95.

W. A. Jordan, "Producer Protection, Prior Market Structure and the Effects of Government Regulation," *J. Law Econ.*, Apr. 1972, *15*, 151–76.

P. L. Joskow, "The Determination of the Allowed Rate of Return in a Formal Regulatory Hearing," *Bell J. Econ.*, Autumn 1972, *3*, 632–44.

______, "Inflation and Environmental Concern: Structural Changes in the Process of Public Utility Price Regulation," *J. Law Econ.*, Oct. 1974, *17*, 291–327.

W. Landes and R. Posner, "The Private Enforcement of Law," *J. Legal Stud.*, Jan. 1975, *5*, 1–46.

L. W. Lee, "Some Models of Antitrust Enforcement," *Southern Econ. J.*, July 1980, *47*, 147–55.

R. Duncan Luce and Howard Raiffa, *Games and Decisions*, New York 1957.

Roger G. Noll, *Reforming Regulation*, Washington 1971.

Bruce M. Owen and Ronald Braeutigam, *The Regulation Game: Strategic Use of the Administrative Process*, Cambridge, Mass. 1978.

S. Peltzman, "Toward a More General Theory of Regulation," *J. Law Econ.*, Aug. 1976, *19*, 211–40.

R. Posner, "Theories of Economic Regulation," *Bell J. Econ.*, Autumn 1974, *5*, 335–58.

John Rawls, *A Theory of Justice*, Cambridge 1971.

Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, Mass. 1960.

G. W. Schwert, "Public Regulation of National Securities Exchanges: A Test of the Capture Hypothesis," *Bell J. Econ.*, Spring 1977, *8*, 128–50.

Amartya K. Sen, *Collective Choice and Social Welfare*, San Francisco 1970.

Martin Shubik, *Strategy and Market Structure: Competition, Oligopoly and the Theory of Games*, New York 1959.

George Stigler, "A Theory of Oligopoly," *J. Polit. Econ.*, Feb. 1964, *72*, 44–59.

______, "The Theory of Economic Regulation," *Bell J. Econ.*, Spring 1971, *2*, 3–21.

______, *The Citizen and the State*, Chicago 1975.

G. Tullock, "The Transitional Gains Trap," *Bell J. Econ.*, Autumn 1975, *6*, 671–78.

Clair Wilcox and William G. Shepherd, *Public Policies Toward Business*, 5th ed., Homewood 1975.

D. M. Winch, "Political Economy and the Economic Policy," *Can. J. Econ.*, Nov. 1977, *10*, 547–64.

E. Zimmerman, "The Legal Framework of Competitive Policies Toward Regulated Industries," in Almarin Phillips, ed., *Promoting Competition in Regulated Markets*, Washington 1975.

Basic Research and Productivity Increase in Manufacturing

By EDWIN MANSFIELD*

Basic research is defined by the National Science Foundation (NSF) as "original investigation for the advancement of scientific knowledge... which do[es] not have immediate commercial objectives" (1959, p. 124). A hotly debated topic among economists, scientists, technologists, and policymakers[1] is: Does basic research, as contrasted with applied research and development, make a significant contribution to an industry's or firm's rate of technological innovation and productivity change? Although economic studies (see Zvi Griliches, 1980; my 1968 book and my 1980 paper; Nestor Terleckyj) indicate that *R&D* expenditures have been directly related to its rate of productivity change, they have been unable to shed light on this question, because no attempt has been made to disaggregate *R&D*. (See my 1977 paper.) One purpose of this paper is to fill this void. Specifically, the paper attempts to determine whether an industry's or firm's rate of productivity change in recent years has been related to the amount of basic research it performed, when other relevant variables (such as its rate of expenditure on applied *R&D*) are held constant.

A second purpose of this paper is to present data obtained from 119 firms concerning past and prospective changes in the composition of their *R&D* expenditures. In recent years, there has been widespread feeling that American industry has reduced the proportion of its *R&D* expenditures going for relatively basic, long-term, and risky projects. However, very few data are available regarding the extent to which such changes in the composition of industrial *R&D* expenditures have taken place. And no information whatsoever seems to be available concerning firms' expectations regarding future changes of this sort. My results seem to be the first data on this subject, about which there is currently so much discussion.

I. Interindustry Differences in the Rate of Productivity Increase: The Model

The first step is to analyze the relationship between an industry's rate of productivity increase and the amount of basic research it performed. The model used here is essentially the same as that employed by Griliches, Terleckyj, and myself, among others, except that research and development is disaggregated into two parts: basic research, and applied research and development. In a particular industry, the production function is assumed to be

$$Q_t = Ae^{\lambda t} R_{tb}^{\alpha 1} R_{ta}^{\alpha 2} L_t^{\nu} K_t^{1-\nu} \tag{1}$$

where Q_t is the industry's value-added in year t, R_{tb} is the industry's stock of basic research capital, R_{ta} is the industry's stock of applied *R&D* capital, L_t is the industry's labor input, and K_t is the industry's stock of physical capital in year t.[2] Thus, the annual rate of change of total factor productivity is

---

*Professor of economics, University of Pennsylvania. This paper is based on work supported by a grant from the National Science Foundation. Of course, the Foundation takes no responsibility for the views expressed here. My thanks go to Z. Griliches, L. Lederman, R. Nelson, and K. Shell for comments on an earlier draft, and to J. Kendrick for providing me with unpublished data.

[1] For two influential studies, see Illinois Institute of Technology Research Institute and Chalmers Sherwin and Raymond Isenson. Of course, the mere fact that industry carries out some basic research does not prove that it makes a significant contribution to innovation and productivity change. Also, see my coauthored book, 1977, and Richard Nelson.

[2] The concept of a stock of research capital has been used by most investigators in this field. (See Griliches 1973 and 1980; my 1968 book; Terleckyj.) Our definitions of R_{tb} and R_{ta} are in accord with those used in previous work. Specifically, the stock of basic research capital in year t equals $\Sigma_i w_{bi} X_{t-i}$, where X_{t-i} is the

$$\rho_t=\lambda+\alpha_1\frac{dR_{tb}/dt}{R_{tb}}+\alpha_2\frac{dR_{ta}/dt}{R_{ta}}$$

Since $\alpha_1\dfrac{dR_{tb}/dt}{R_{tb}}=\dfrac{\partial Q_t}{\partial R_{tb}}\cdot\dfrac{dR_{tb}/dt}{Q_t}$

and $\alpha_2\dfrac{dR_{ta}/dt}{R_{ta}}=\dfrac{\partial Q_t}{\partial R_{ta}}\cdot\dfrac{dR_{ta}/dt}{Q_t}$

$$\rho_t=\lambda+\phi_0\frac{dR_{tb}/dt}{Q_t}+\phi_1\frac{dR_{ta}/dt}{Q_t} \tag{2}$$

where $\phi_0=\partial Q_t/\partial R_{tb}$ and $\phi_1=\partial Q_t/\ R_{ta}$.

Because the rate of growth of each industry's stock of *R&D* capital (both basic and applied) cannot be measured directly, I assume, as Griliches and Terleckyj did, that an industry's expenditure on *R&D* during year t is approximately equal to that year's change in the industry's stock of *R&D* capital. For this assumption to hold, the depreciation of the *R&D* capital and the lag in the effect of this investment must be small enough to be ignored.[3] Although the assumption of no lags is a useful starting point, I relax this assumption later. (See fn. 7.) According to John Kendrick, Terleckyj, and others, an industry's annual rate of change of total factor productivity also depends on the extent to which the industry's workers are unionized.[4] To avoid specification errors, I therefore include u_t, the percent of the industry's workers that are union members, as an additional independent variable in equation (2). Thus, the resulting equation is

$$\rho_t=\lambda+\phi_0\frac{X_t}{Q_t}+\phi_1\frac{A_t}{Q_t}+\phi_2u_t \tag{3}$$

where X_t is the industry's expenditure on basic research in year t, and A_t is the industry's expenditure on applied *R&D* in year t.

During the period to which my data pertain, there seems to have been only a limited amount of variation in an industry's value of X_t/Q_t, A_t/Q_t, and u_t. Thus, as a first approximation, it seems reasonable to suppose that $X_t/Q_t=B$, $A_t/Q_t=D$, and $u_t=u$. Then, if an industry's values of λ, ϕ_0, ϕ_1, and ϕ_2 are statistically independent of its values of B, D, and u, it follows that

$$\rho_i=\bar{\lambda}+\bar{\phi}_0B_i+\bar{\phi}_1D_i+\bar{\phi}_2u_i+z_i \tag{4}$$

where ρ_i is the annual rate of change of total factor productivity in the ith industry during this period; B_i, D_i, and u_i are the values of B, D, and u in the ith industry, $\bar{\lambda}$, $\bar{\phi}_0$, $\bar{\phi}_1$, and $\bar{\phi}_2$ are the average values of λ, ϕ_0, ϕ_1, and ϕ_2 in all relevant industries, and z_i is a random error term.[5]

II. Econometric Results

To what extent can equation (4) explain the differences among manufacturing industries in the rate of increase of total factor productivity? To answer this question, we must have data concerning ρ_i, B_i, D_i, and u_i. Kendrick has provided data concerning ρ_i for the twenty 2-digit SIC manufacturing industries during 1948–66. Following Terleckyj, who has used a similar sort of model to analyze Kendrick's data, each of the independent variables (other

---

expenditure by the industry on basic research in year $t-i$, and w_{bi} is the proportion of this expenditure that is still productive in year t. The stock of applied *R&D* capital in year t equals $\Sigma_iw_{ai}A_{t-i}$, where A_{t-i} is the expenditure by the industry on applied *R&D* in year $t-i$, and w_{ai} is the proportion of this expenditure that is still productive in year t.

[3] The rationale given by Griliches and Terleckyj for ignoring depreciation is that, because so little was spent on *R&D* prior to World War II and because the rate of growth of *R&D* expenditures in the 1950's was so high, the amount of depreciation that would have been able to build up during the period covered by our data should be small. Relatively little is known about the length of the relevant lags, although some data on this score are presented in my 1968 book and my 1980 paper.

[4] See Kendrick and Terleckyj. Another variable that Terleckyj found to be statistically significant in some specifications was the percent of the industry's sales to nongovernment buyers. I included this variable in those equations where he found it was significant. For the results, see fn. a of Table 1.

[5] With regard to A_t/Q_t and X_t/Q_t, the earliest comparable data are for 1953. Comparing an industry's value of A_t/Q_t in 1953 with that in 1966, I find that there is a high correlation ($r=.91$) between them, and that the intertemporal variation in A_t/Q_t is much smaller than the interindustry variation in it. Similar results were found for X_t/Q_t.

TABLE 1—RESULTS OF REGRESSION TO EXPLAIN RATES OF PRODUCTIVITY INCREASE IN TWENTY MANUFACTURING INDUSTRIES, 1948–66[a]

	Independent Variables[b]								
Constant	B_i	D_i	U_i	D_{pi}	D_{gi}	I_i	I_{pi}	Y_i	$\bar{R}^2$
4.80	1.49	.068	−.056	–	–	–	–	–	.63
	(3.34)	(2.75)	(3.90)						
4.74	1.33	–	−.055	.105	.055	–	–	–	.63
	(1.51)		(3.54)	(0.59)	(0.80)				
4.37	1.66	−.084	−.055	–	–	.54	–	–	.71
	(4.55)	(1.59)	(4.73)			(3.10)			
4.28	1.50	−.18	−.054	–	–	–	.74	–	.80
	(4.94)	(0.71)	(5.57)				(4.41)		
4.46	.285	.084	−.061	–	–	–	–	.065	.86
	(0.81)	(5.17)	(6.37)					(4.80)	
4.83	.682	–	−.064	−.023	.122	–	–	0.68	.87
	(1.27)		(6.32)	(0.21)	(2.88)			(4.90)	
4.46	.572	.005	−.059	–	–	.27	–	.054	.86
	(1.67)	(.14)	(7.02)			(2.07)		(4.16)	
4.42	.665	.035	−.058	–	–	–	.37	.045	.86
	(1.85)	(1.29)	(6.81)				(2.11)	(2.90)	

[a] Besides these regressions, I calculated all of Terleckyj's equations in his Tables 4–9 with B_i and Y_i added (and with the nonsignificant non-*R&D* variables in his equations omitted). In practically all cases, the regression coefficient of B_i is significant when it alone is added. When both B_i and Y_i are added, the regression coefficient of Y_i is always highly significant, but the regression coefficient of B_i is significant in only one-third of the equations. The estimated coefficients of the independent variables are generally similar to those shown here.

[b] All of the independent variables, as well as the dependent variable, are expressed as percentages. The number in parentheses below each regression coefficient is its *t*-value.

than u_i, which pertains to 1953) is measured as of 1958. To measure B_i, I calculated the ratio of each industry's expenditure on basic research to its value-added in 1958, based on data published by the National Science Foundation and the Commerce Department.[6] Terleckyj has provided data concerning D_i and u_i. Using these data, I regressed ρ_i on B_i, D_i, and u_i, with the results shown in Table 1. All three independent variables turn out to be highly significant, and over 60 percent of the variation among industries in the rates of productivity increase are explained by the equation.[7]

---

[6] The data regarding B_i are based on the basic research expenditures reported in the 1961 edition of *Research and Development in Industry*. Note that both government- and company-financed basic research are included. In two cases, the NSF provided data regarding two of the industies combined. I allocated the basic research expenditures between the two industries in proportion to their applied *R&D* expenditures (as reported in Terleckyj). This procedure is very unlikely to result in any serious error, since these industries do very little basic research. If an industry did no applied *R&D* in 1958, according to Terleckyj, and the NSF provided no indication that it carried out any basic research, I assumed that its basic research expenditures were zero in 1958.

[7] The effects of *R&D* often occur with a lag, and the lag for basic research is generally thought to be much longer than for applied *R&D*. Because so little is known about the length of these lags, previous work along this line has sometimes (but not always) ignored them; but once we separate basic research from applied *R&D*, it is hazardous to continue this procedure, since the lags in the effects of basic research may be quite long. If an expenditure on basic research is included in the stock of basic research capital only after a lag of θ_1 years, and if an expenditure on applied *R&D* is included in the stock of applied *R&D* capital only after θ_2 years, equation (3) must be modified as follows:

$$\rho_t=\lambda+\phi_0\frac{X_{t-\theta_1}}{Q_t}+\phi_1\frac{A_{t-\theta_2}}{Q_t}+\phi_2u_t$$

If r_1 is the average annual rate of growth of the industry's value-added between year $t-\theta_1$ and year t, and if r_2 is the average annual rate of growth of the industry's value-added between year $t-\theta_2$ and year t, it follows that

$$\rho_t=\lambda+\phi_3\frac{X_{t-\theta_1}}{Q_{t-\theta_1}}+\phi_4\frac{A_{t-\theta_2}}{Q_{t-\theta_2}}+\phi_2u_t$$

(over)

To see how robust these results are, I experimented with some other specifications of the regression equation. First, I divided the total amount of applied *R&D* into two parts: the amount that is privately financed, and the amount that is financed by the government. This means that two independent variables, D_{pi} (the amount of privately financed applied *R&D* divided by value-added) and D_{Gi} (the amount of government-financed applied *R&D* divided by value-added), must be substituted for D_i. As shown in Table 1, this substitution has relatively little effect on the estimate of $\bar{\phi}_0$, although it reduces its statistical significance.

Second, I added to equation (3) a fourth independent variable I_i, which measures the amount of *R&D* that is embodied in an industry's purchased inputs (divided by its value-added). Terleckyj has provided estimates of this variable for these twenty industries. As shown in Table 1, when this variable is added, it is highly significant, but its inclusion has little effect on the size or statistical significance of the estimate of $\bar{\phi}_0$. Similarly, if this variable is defined to include only privately financed *R&D*, the results are essentially the same, as shown in Table 1 (where this variable is denoted by I_{pi}).

These findings seem to indicate a strong relationship between the amount of basic research carried out by an industry and the industry's rate of productivity increase during 1948–66. However, one wonders whether, since the distinction between basic research and applied research is often nebulous, this relationship may not reflect the fact that industries that carry out relatively large amounts of long-term *R&D* tend to have relatively high rates of productivity increase. In other words, basic research may be a proxy for long-term *R&D*. To shed some light on this question, I added another variable, Y_i (the percent of the industry's firms that expected their *R&D* expenditures in 1958 to pay off in no less than six years), to the four regressions discussed above.[8] As shown in Table 1, the results suggest that much of the apparent effect of basic research may really be due to long-term *R&D*. When this additional variable is introduced, its effect is highly significant in every case,

---

where $\phi_3 = \phi_0 \div (1+r_1)^{\theta 1}$ and $\phi_4 = \phi_1 \div (1+r_2)^{\theta 2}$. If the independent variables in this equation remain relatively constant in the relevant period, and if an industry's values of λ, ϕ_2, ϕ_3, and ϕ_4 are statistically independent of the values of these independent variables,

$$\rho_i = \bar{\lambda} + \bar{\phi}_3 B_i'' + \bar{\phi}_4 D_i'' + \bar{\phi}_2 u_i + z_i$$

where B_i'' is the value of $X_{t-\theta 1}/Q_{t-\theta 1}$ and D_i'' is the value of $A_{t-\theta 2}/Q_{t-\theta 2}$ in the ith industry during the relevant period. Although we do not know θ_1 and θ_2 with much accuracy, it is very likely that there is a high correlation between B_i'' and B_i, as well as between D_i'' and D_i. For as far back as reasonably accurate data exist, industries with relatively high ratios of basic research to value-added in one year have tended to have relatively high ratios in other years, and industries with relatively high ratios of applied *R&D* to value-added in one year have tended to have relatively high ratios in other years. The coefficient of correlation between an industry's ratio of applied *R&D* expenditures to value-added in 1953 and its ratio in 1975 is 0.93. The coefficient of correlation between an industry's ratio of basic research expenditures to value-added in 1953 and its ratio in 1975 is 0.85. Even if the value of θ_1 or θ_2 varies from industry to industry, there is likely to be a high correlation between B_i'' and B_i and between D_i'' and D_i. Thus, B_i and D_i can be regarded as proxies for B_i'' and D_i''. Viewed in this way, my regression equations, since they are based on these proxies rather than on B_i'' and D_i'', cannot be used to estimate the rates of return from basic research or from applied *R&D*, unless one can estimate the unknown values of θ_1 and θ_2, as well as the relationship between these proxies and the variables for which they stand. However, the results are useful in testing whether ϕ_0 equals zero, since, if this is the case, the regression coefficient of B_i in the regressions should be zero. Thus, despite the presence of these lags, my results are useful in testing whether an industry's or firm's rate of productivity change has been affected by the amount of basic research it performed, when its rate of expenditure on applied *R&D* is held constant.

[8] Data concerning Y_i were obtained from McGraw Hill's *Business' Plans for New Plants and Equipment, 1958–1961*, p. 8. The figure for primary metals is the average of that given for "iron and steel" and for "nonferrous metals." The figure for transportation equipment is the average of that given for "autos" and that given for "transportation equipment." The figure for fabricated metal products and instruments is that given for "other metalworking." Since three of the industries carried out no applied *R&D* expenditures in 1958 according to Terleckyj, no data regarding Y_i exist for these industries.

whereas the effect of B_i becomes smaller and often statistically nonsignificant.

In addition, to extend the time period to which the data pertain, I used unpublished data by Kendrick to calculate ρ_i for these twenty manufacturing industries during 1948–76. When these values of ρ_i are used in Table 1, the magnitude of the regression coefficient of B_i—and the number of specifications where it is statistically significant—is about the same as in Table 1. However, the regression equations do not fit nearly as well, probably due in part to the increased importance during 1966–76 of regulatory considerations and a variety of other factors not included in the model. Also, the regression coefficients of Y_i are no longer statistically significant. This is not surprising because there have been considerable changes since 1958 (the year to which Y_i pertains) in the extent to which various industries have focused their *R&D* on long-term work. Much more will be said on this score in Section IV.[9]

III. Interfirm Differences in the Rate of Productivity Increase

Besides investigating the relationship between basic research and productivity increase at the industry level, I also investigated this relationship at the firm level. Once again, let us assume that equation (1) holds, but now Q_t is the firm's valued-added in year t, R_{tb} is the firm's stock of basic research capital, R_{ta} is the firm's stock of applied *R&D* capital, and so forth. Making the same assumptions as in Section I, it follows that

$$(5) \qquad \rho_i' = \bar{\lambda}' + \bar{\phi}_0' B_i' + \bar{\phi}_1' D_i' + z_i'$$

where ρ_i' is the annual rate of change of total factor productivity in the ith firm; B_i' and D_i' are the ratio of basic research expenditure to value-added and the ratio of applied *R&D* expenditure to value-added in the ith firm; and $\bar{\lambda}'$, ϕ_0', ϕ_1', and z_i' are analogous (at the firm level) to $\bar{\lambda}$, ϕ_0, ϕ_1, and z_i. Equation (4) is analogous to equation (3) except that u_i is excluded because it has the same value for all firms in the industry.[10]

To see how well this model fits, and to estimate ϕ_0' and ϕ_1', I obtained data concerning value-added and employment during 1960–76, and data concerning investment in plant and equipment during 1920–76 for ten major petroleum firms and for six major chemical firms.[11] The stock of physical capital was estimated annually for each firm during 1960–76. To obtain such an estimate for a particular year, the investment data for each firm for previous years were deflated, depreciated (using the Jack Faucett-Bureau of Labor Statistics (BLS) depreciation rates for plant and equipment in chemicals and petroleum refining), and summed. Using these data concerning each firm's stock of physical capital each year, together with the data concerning each firm's (deflated) value-added and employment each year, I estimated the annual rate of total factor productivity change for each firm during 1960–76.[12] In addition to these estimates of ρ_i', we need estimates of B_i' and D_i' for each of these firms. As a by-product of previous research, I had these data for all of these firms for 1964 or for 1965.[13]

---

[9] Kendrick kindly provided me with estimates of total factor productivity in each industry during 1948–76. These estimates will appear in his forthcoming book. I calculated the annual percentage rate of growth of total factor productivity between 1948 and the year since 1966 when total factor productivity reached a maximum. (If total factor productivity is less than in a previous year, the industry is probably not operating on its production function, and the inclusion of such an observation may result in misleading results.) The year when it reaches a maximum is almost always 1976 or 1973. When the resulting values of ρ_i are inserted in the regressions in Table 1, the regression coefficient of B_i is statistically significant at the .05 level in about the same proportion of cases as in Table 1.

[10] Of course, I could have included a variable measuring interfirm differences in the extent of unionization, but this did not seem worthwhile, because the interfirm differences in this regard seemed small.

[11] The data concerning value-added and employment were obtained from *Moody's Industrial Manual* and from the annual reports of the firms. The figures regarding value-added were deflated by the Bureau of Economic Analysis's implicit deflator for the industry. Most of the data concerning investment in plant and equipment come from annual reports and *Moody's*. However, for the earliest years, it was often necessary to estimate a firm's investment in plant and equipment during certain years by assuming it was proportional to the change in the firm's total assets. Although these estimates for the early years are crude, the results should be influenced very little by whatever errors they contain, since the investment figures for the early years have relatively little weight in the capital stock figures. This is because most of the plant and equipment acquired in the early years were depreciated by 1960. In the case of the petroleum firms, the analysis was carried only up to 1971, not 1976, because of the gyrations in oil prices and the unsettled conditions in the industry in recent years.

To estimate the parameters of equation (5), I regressed each firm's value of ρ_1' on its values of B_i' and D_i', the result being

$$(6)\quad \rho_i' = \underset{(7.51)}{2.15} + \underset{(2.08)}{1.78B_i'} + \underset{(0.86)}{0.10D_i'}$$

$$\bar{R}^2 = 0.58$$

where the t-statistics are shown in parentheses below the regression coefficients, and where both the dependent and independent variables are expressed as percentages. To determine whether there were interindustry differences between the chemical and petroleum firms, I introduced as an additional independent variable a dummy variable that equals 1 if the firm is in the chemical industry and 0 if it is in the petroleum industry. This dummy variable was not statistically significant, so it was omitted from equation (6).[14]

The results are surprisingly similar to those obtained at the industry level. There is a statistically significant tendency for a firm's rate of productivity increase during 1960–76 to be directly related to how much basic research it carried out in 1964, when the amount spent on applied *R&D* is held constant. Moreover, the regression coefficient of B_i' in equation (6) is not too different from the regression coefficient of B_i in equation (4).

Previous studies (for example, Griliches, 1980) have used such regression coefficients as estimates of the private rate of return from *R&D*. For well-known reasons, some of which are discussed in footnote 7, estimates of this sort should be treated with considerable caution. Nonetheless, it is interesting to see how an estimate of this sort based on these data compares with previous estimates, since so little is known about the changes that have occurred in recent years in the profitability of *R&D*. (Practically all previous work of this kind pertains to the early 1960's.) To obtain such an estimate, I regress ρ_i' on $(B_i' + D_i')$, the result being

$$(7)\quad \rho_i' = \underset{(7.06)}{2.2} + \underset{(4.08)}{.275}\,(B_i' + D_i')$$

$$\bar{R}^2 = .49$$

which implies that the rate of return from all *R&D* (basic and applied) was about 27 percent in 1960–76 in these chemical and petroleum firms. This estimate is lower than corresponding estimates (using the same technique and the same industries) for the early 1960's, a result that is consistent with the reduced rate of growth of *R&D* expenditures by industry during the early 1970's and with opinions expressed by many *R&D* managers.[15]

---

[12] The investment figures were deflated by price deflators provided by the BLS. Then the Faucett-BLS depreciation rates were applied, and the results were summed to obtain the firm's capital stock each year. Then, for each firm, I computed $g_t = lnQ_t - b lnL_t - (1 - b) lnK_t$, where Q_t is the firm's deflated value-added in year t, L_t is its employment in year t, K_t is its capital stock in year t, and b is its average proportion of value-added going for wages during 1960–76. Finally, I regressed g_t on t, the resulting regression coefficient being the estimate of ρ_i'. Years when the firm's value-added was less than in some previous year were omitted because the firm was apparently operating well below its capacity. Clearly, in an analysis of this sort, the data should apply to periods when the firm is operating reasonably close to capacity. The data described in this and the previous footnote were obtained, and the calculations were made, by Nancy Fox as part of her work toward a doctoral dissertation at the University of Pennsylvania.

[13] Some of these data were described in my coauthored book, 1971. The rest come from an unpublished survey carried out by Henry Armour and David Teece. To check whether the ratio of applied *R&D* expenditures to value-added remained relatively constant during the relevant period, I compared the values of this ratio in 1976 and (about) 1964 for the six chemical firms. In most firms, the 1976 ratio was within 25 percent of the 1964 ratio. The interyear differences tend to be much smaller than the interfirm differences. There was a high correlation between a firm's 1976 ratio and its 1964 ratio. With regard to the ratio of basic research expenditures to value-added, the results are much the same.

[14] The same result was obtained in my 1968 book on the basis of earlier data.

IV. Recent Changes in the Composition of Industrial *R&D* Expenditures

In recent years, there has been a widespread feeling that industry has been devoting a smaller share of its *R&D* expenditures to basic research, long-term projects, and risky and ambitious projects. Unfortunately, however, little or no data have been available on this score. To make some progress toward filling this gap, I obtained information from 119 firms concerning the changes that have occurred in this regard between 1967 and 1977, and changes they expect between 1977 and 1980. The firms included in my sample, all of which spent over $10 million on *R&D* in 1976, accounted for about one-half of all industrial *R&D* expenditures in the United States in 1976. Although this sample has its shortcomings, it is certainly among the most comprehensive surveys of this type ever carried out.[16]

The results presented in Table 2 indicate that the proportion of *R&D* expenditures devoted to basic research declined between 1967 and 1977 in practically every industry. In the aerospace, metals, electrical equipment, office equipment and computer, chemical, drug, and rubber industries, this proportion dropped substantially. In the sample as a whole, the proportion fell about one-fourth, from 5.6 percent in 1967 to 4.1 percent in 1977. According to the firms' forecasts, there is no evidence that this drop will continue during 1977–80, but neither is there any evidence that the proportion will rise very much. For the sample as a whole, the proportion is expected to be about 4.3 percent in 1980.[17]

In four-fifths of the industries, there was also a decline between 1967 and 1977 in the proportion of *R&D* expenditures devoted to relatively risky projects (specifically, ones with less than a fifty-fifty estimated chance of success). In some industries, like metals, chemicals, aircraft, drugs, and rubber, this reduction has been rather large. Between 1977 and 1980, there will be some increase overall in the proportion devoted to relatively risky projects, according to the firm's forecasts, with the result that the average proportion for the sample as a whole is expected to get closer to its 1967 level.

Despite the decrease in the proportion of *R&D* expenditures devoted to basic research, the proportion devoted to relatively long-term projects (specifically, those lasting five or more years) did not decline appreciably between 1967 and 1977 in the sample as a whole. In some industries, like aircraft,

---

[15] In 1960, the estimated rate of return in these industries averaged about 30 percent or more (see my 1968 book). In the early 1960's, it was about 40 percent, according to Griliches, 1980. Of course, all of these estimates neglect lags, and should be treated with considerable caution.

[16] Data were obtained (through correspondence or interviews) from practically all of the major chemical firms, from about three-quarters of the major aerospace, automotive, petroleum, metals, and office equipment and computer firms, from about two-thirds of the major rubber, soap, and machinery firms, from about one-half of the major drug, food, and instruments firms, and from about one-quarter of the major electronics and electrical equipment firms. By a major firm, I mean one that spent over $10 million on *R&D* in 1976, according to the survey in *Business Week*, June 27, 1977. The data were obtained in practically all cases from the firm's vice president of research and development. One obvious limitation of the sample is that it focuses entirely on large firms. For an interesting recent study which deals with some of the topics discussed here, see Howard Nason et al.

[17] According to NSF's figures, 4.7 percent of company-financed *R&D* expenditures in manufacturing went for basic research in 1975 (the most recent year for which NSF has published data). This is reasonably close to my 1977 figure of 4.1 percent. In 1968, NSF put this percentage at 6.9 percent, whereas my 1967 figure is 5.6 percent. For nine industries in 1977 and seven industries in 1967, I could compare my percentage with NSF's percentage. I found that my percentages are positively correlated with NSF's, but that they do not always agree closely. This is not surprising, since my definition of each of these industries differs somewhat from NSF's and NSF's coverage is more comprehensive than mine. Nonetheless, there is close agreement with regard to the change in this percentage in all manufacturing. NSF's data indicate a decrease of 32 percent in this percentage between 1968 and 1975, and my data indicate a decrease of 27 percent between 1967 and 1977. For the NSF data, see *Research and Development in Industry*, 1975 and 1978.

TABLE 2—PERCENTAGE OF COMPANY-FINANCED *R&D* EXPENDITURE GOING FOR BASIC, LONG-TERM AND RELATIVELY RISKY PROJECTS: 1967; 1977; 1980[a]

	Basic Research			Projects Lasting Five or More Years			Projects Aimed at Entirely New Products and Processes			Projects with Less than 50-50 Estimated Chance of Success		
Industry	1967	1977	1980	1967	1977	1980	1967	1977	1980	1967	1977	1980
Metals	6.2	2.4	2.4	26	22	22	28	18	19	18	11	12
Chemicals	7.3	5.9	5.8	43	39	39	37	33	31	37	30	27
Aerospace	8.5	1.7	1.7	31	24	24	35	26	28	35	27	29
Automobiles	0.3	0.2	0.3	18	20	23	15	16	–	47	45	45
Petroleum	9.1	8.1	9.7	26	27	29	21	25	26	24	23	24
Drugs	20.7	16.4	15.3	63	66	63	76	68	65	46	40	38
Food	2.6	4.1	4.3	16	16	22	18	22	25	19	18	20
Instruments	3.7	3.3	3.3	40	40	41	24	22	22	11	9	10
Soap and Cosmetics	3.4	3.9	3.6	23	30	37	38	48	55	43	55	55
Machinery	4.2	3.6	3.4	34	31	32	23	18	20	8	5	5
Electronics and Electrical Equipment	6.4	4.6	4.8	10	14	14	37	45	47	3	5	5
Office Equipment and Computers	3.5	1.8	–	55	55	–	60	60	–	9	9	–
Stone, Clay and Glass	8.0	7.1	8.1	35	43	38	37	28	27	28	24	21
Rubber	2.7	1.7	1.7	9	5	4	32	28	28	37	27	22
Miscellaneous	5.1	2.5	3.5	30	27	27	21	22	27	18	13	17
Total	5.6	4.1	4.3	34	34	–	36	34	34	28	25	27

[a] The 1980 figures are the firms' forecasts. Some 1980 figures are not entirely comparable with earlier ones because some firms did not provide 1980 forecasts; in cases where the lack of comparability is serious, no 1980 figure is given.

chemicals, metals, and rubber, there was a substantial decline; but in other industries, like drugs, there was an increase in this proportion. In most industries, the firms forecast that the proportion of *R&D* expenditures devoted to relatively long-term projects will increase somewhat between 1977 and 1980.[18] The proportion of *R&D* expenditures aimed at entirely new products and processes (rather than improvements and modifications of existing products and processes) declined somewhat between 1967 and 1977.[19] Again, there are marked differences among industries in the amount and direction of change. Between 1977 and 1980, little change was expected to occur in this proportion in most industries.

A multiple regression analysis provides no evidence of any relationship between a firm's size, profitability, or percentage of *R&D* expenditures devoted to sales, on the one hand, and the change in the proportion of the firm's *R&D* expenditures devoted to basic research between 1967 and 1977 (or expected to occur between 1977 and 1980), on the other. Moreover, there is no relationship between these three variables and the change in the proportion of the firm's *R&D* expenditures devoted to relatively risky projects, or to relatively long-term projects, or to projects aimed at entirely new products and processes.[20] Further, there is only a very weak relationship, if any, between the change in the proportion of a firm's *R&D* expenditures devoted to basic research, and the change in the proportion devoted to relatively risky projects, relatively long-term projects, or projects aimed at entirely new products or processes.[21]

---

[18] For other data regarding the proportion of firms' *R&D* expenditures that goes for relatively long-term projects, see my coauthored book, 1971, and my 1980 paper. These data are quite consistent with the results in Table 2.

[19] The distinction between an entirely new product or process and an improved or modified product or process is often arbitrary (although it is often used). So long as each respondent uses the same definition for one year as for another year, the results should help to portray the changes that have occurred over time in a given industry. But a comparison in this regard of one industry with another may be misleading since the definitions used in one industry may be quite different from those used in another industry.

Besides obtaining data from these 119 firms, I also asked those that have cut back on the proportion of their *R&D* expenditures going for basic research and relatively risky and long-term projects why such a change has occurred. The reason most frequently given was the increase in government regulations, which has reduced the profitability of more fundamental and longer-term projects. This reason was advanced particularly often by the chemical and drug firms.[22] Another reason advanced by some of the respondents was that breakthroughs are more difficult to achieve than in the past, because the field has been more thoroughly worked over. Also, high rates of inflation were cited.

V. Conclusions

My results indicate that there is a statistically significant and direct relationship between the amount of basic research carried out by an industry or firm and its rate of increase of total factor productivity, when its expenditures on applied *R&D* are held constant. To some extent, this may reflect a tendency for basic research findings to be exploited more fully by the industries and firms that were responsible for them. Or it may reflect a tendency for applied *R&D* to be more effective when carried out in conjunction with some basic research. (See William Price and Lawrence Bass.) Whether the relevant distinction is between basic and applied research is by no means clear; there is some evidence that basic research may be acting to some extent as a proxy for long-term *R&D*. Holding constant the amount spent on both applied *R&D* and basic research, an industry's rate of productivity increase during 1948–66 seems to be directly and significantly related to the extent to which its *R&D* was long term. This appears to be the first systematic evidence that the composition, as well as the size, of an industry's or firm's *R&D* expenditures affects its rate of productivity increase.

My findings also indicate that the composition of many industries' *R&D* expenditures has changed in the last decade. Practically all industries have cut the proportion of their *R&D* expenditures going for basic research. Most industries have cut the proportion going for relatively risky projects. Some, but not most, industries have cut the proportion going for relatively long-term projects.[23] Particularly in the chemical, aerospace, and rubber industries, it appears that such changes may put a damper on the rate of productivity increase.[24] But without much more accurate data concerning the lags in the effects of *R&D*, it is difficult to say how quickly these effects will be felt.

---

[20] I ran four regressions, the dependent variables being (a) the change between 1967 and 1977 in the proportion of a firm's *R&D* expenditures devoted to basic research, (b) the change between 1967 and 1977 in the proportion devoted to projects where there was less than a 50-50 estimated chance of success, (c) the change between 1967 and 1977 in the proportion devoted to projects lasting five or more years, and (d) the change between 1967 and 1977 in the proportion aimed at entirely new products or processes. Each of these dependent variables was regressed on the firm's sales in 1976, its ratio of *R&D* expenditures to sales in 1976, its ratio of profits to net worth in 1976, and industry dummy variables. Other than some of the dummy variables, none of the independent variables in any of these multiple regressions was statistically significant.

[21] I regressed three dependent variables—the change between 1967 and 1977 in the proportion of a firm's *R&D* expenditures devoted to projects lasting five or more years, the change in the proportion aimed at entirely new products or processes, and the change in the proportion devoted to projects where there was less than a 50-50 estimated chance of success—on the change between 1967 and 1977 in the proportion devoted to basic research. Although the sample size exceeded 100, in only one case was the regression statistically significant, and in this case, r^2 was only .06.

[22] About 80 percent of the relevant chemical firms and practically all of the relevant drug firms cited this reason.

[23] To some extent, this may be misleading if the length of projects is increased by regulatory and other such factors. In this case, the proportion spent on what formerly would have been regarded as long-term projects may have declined, even though this does not show up in the data because of the general lengthening of projects. The drug industry is an example where regulatory factors seem to have prompted a general lengthening of projects.

In conclusion, the limitations of this study should be noted. In particular, the model adopted in Section I is highly simplified, and the data in Sections II and III often are rough. Even in Section IV, where the data pertain to about one-half of all *R&D* by *U.S.* manufacturing companies, there are bound to be sampling errors, since not all firms are included. For these and other reasons, my findings must be regarded as tentative. Much more work is needed in this area. For example, it is important that we try to determine the extent to which the observed relationship between the composition of an industry's or firm's *R&D* expenditures and its rate of productivity change is due to the influence of other factors on both variables.[25] Nonetheless, despite its limitations, I believe that this study is of considerable interest, since it is the first attempt to disaggregate *R&D* expenditures (which are notoriously heterogeneous) and to investigate the effects of the composition of an industry's or firm's *R&D* expenditures on its rate of productivity increase.

---

[24] Besides this damper due to changes in the composition of *R&D*, there may be a damper on the rate of productivity increase in these industries because of the decrease during the last decade in the ratio of basic research expenditures to value-added and in the ratio of applied *R&D* expenditures to value-added.

[25] It is possible that industries and firms with high rates of productivity growth tend to spend relatively large amounts on basic research, but that their high rates of productivity growth are not due to these expenditures. Much more work is required before one can be entirely confident of the way in which this observed relationship should be interpreted.

REFERENCES

K. Arrow, "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Universities-Nat. Bur. Econ. Res. conference series, Princeton 1962.

Z. Griliches, "Research Expenditures and Growth Accounting," in B. R. Williams, ed., *Science and Technology in Economic Growth*, London 1973.

______, "Returns to Research and Development Expenditures in the Private Sector," in John W. Kendrick and Beatrice N. Vaccara, eds., *New Developments in Productivity Measurement*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, New York 1980.

John Kendrick, *Post War Productivity Trends in the United States, 1948–69*, New York 1973.

______, *Productivity in the United States*, forthcoming.

Edwin Mansfield, *Industrial Research and Technology Innovation*, New York 1968.

______, "Comments on Returns to Research and Development Expenditures," in John W. Kendrick and Beatrice N. Vaccara, eds., *New Developments in Productivity Measurement*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, New York 1980.

______, "Research and Development, Productivity Change, and Public Policy," *Colloquium on Research and Development and Productivity*, National Science Foundation 1977.

______, "Technology and Productivity in the United States," Conference on Postwar Changes in American Economy, Nat. Bur. Econ. Res. 1980.

______ **et al.,** *Research and Innovation in the Modern Corporation*, New York 1971.

______ **et al.,** *The Production and Application of New Industrial Technology*, New York 1977.

Howard Nason, Joseph Steger, and G. Manners, *Support of Basic Research by Industry*, Washington 1978.

R. Nelson, "The Simple Economics of Basic Scientific Research," *J. Polit. Econ.*, June 1959, *67*, 297–306.

W. Price and L. Bass, "Scientific Research and the Innovative Process," *Science*, May 16, 1969, *169*, 802–06.

C. Sherwin and R. Isenson, *First Interim Report on Project Hindsight*, Office of Defense Research and Engineering, Washington 1966.

Nestor Terleckyj, *Effects of R and D on the Productivity Growth of Industries*, Washington 1974.

Illinois Institute of Technology Research Institute, *Technology in Retrospect and Critical Events in Science*, rept. to the National Science Foundation, Washington 1969.

McGraw-Hill Economics Department, *Business Plans for New Plants and Equipment, 1958–1961*, New York 1962.

Moody's Industrial Manual, New York, various issues.

National Science Foundation, *Methodology of Statistics on Research and Development*, Washington 1959.

——, *Research and Development in Industry*, Washington, various issues.

Consensus Beliefs, Information Acquisition, and Market Information Efficiency

By ROBERT E. VERRECCHIA*

In an attempt to provide a precise analytical interpretation for the notion of "information efficiency" introduced by Eugene Fama (1970), Mark Rubinstein suggests that a necessary and sufficient condition for an individual to perceive all his information fully reflected in prices is that he have "consensus beliefs." That is, consensus beliefs are those beliefs which, if held by all individuals in an otherwise similar economy, would generate the same equilibrium prices as in the actual heterogeneous economy. However, the notion of information efficiency embodied in Rubinstein's discussion has been more broadly interpreted as follows.[1] A market is efficient with respect to an information set **A**, say, if the prices it generates are identical to those generated in an otherwise identical economy in which the set **A** accurately describes the information available to each and every market participant; the common, or homogeneous, belief induced by knowledge of **A** is the consensus belief.[2]

If the set of information **A** happens to be the union of the information sets privately available to each and every market participant, then efficiency with respect to **A** can be characterized as a situation in which the consensus belief is at least as correct an assessment (on the basis of **A**) as the assessment privately held by any single investor. Stated more informally, all investors perceive that no less information is fully reflected in prices than that which each individual investor privately knows. With regard to welfare implications, when the consensus belief is at least as correct, investors face a type of fair game in the market. That is, let "excess returns" be defined as the difference in the expected utility some well-informed investor achieves by trading at a price which reflects information inferior to his own vis-à-vis the expected utility he achieves when price reflects information at least as good. Then a price reflecting no less information than what is privately known prohibits excess returns. In effect, prices provide an effective barrier to the exploitation of less well-informed investors by better-informed investors, because no investor can exploit his superior information (in the classical price-taking setting usually attributed to a securities market) whenever a price exists which reflects information which is in no way inferior to his own.[3]

---

*Graduate School of Business, University of Chicago.

[1] See, for example, Avraham Beja, William Beaver, and my 1979b paper.

[2] For example, suppose that a market consisted of two investors, the first of whom had as his sole source of information a set **O**, say, which contained only optimistic bits and pieces of data about a security, and which consequently induced him to regard the future return of the security as "abnormally high." Alternatively, the second investor had as his sole source of information a set **P**, say, which contained only pessimistic bits and pieces of data about the security, and which consequently induced him to regard the future return of the security as "abnormally low." Suppose further that trading against one another in the market resulted in a price that appeared to reflect the fact that the future return for the security would be "average"; this is a natural supposition if both traders have equal weight in the market with regard to the supply and demand of the security. Now if in an otherwise identical market the set of information **A**, where **A** is the union of **O** and **P**, is substituted for both traders' original private sources, **O** and **P**, respectively, and if this knowledge induces them to regard the future return of the security as average, then this will likely cause no price change since prices have already averaged out their diverse opinions. Thus the market for the security will be efficient with respect to **A**, where **A** is the union of **O** and **P**.

[3] To illustrate this, consider the situation described in fn. 2 in which one investor knows **O**, and another knows **P**. Suppose that a third investor enters the market who happens to know **A**, the union of **O** and **P**. Clearly, the third investor is "better informed" than either of the other two, where it will be assumed

In brief, the consensus belief concept is an intuitively plausible definition of market information efficiency which, unlike some of its competitors, lends itself to formal analyses. The purpose of this paper is to consider its robustness. For example, it is well known that prices are determined in part by a type of geometric averaging of each investor's assessment weighted by his tolerance for risk. Therefore, efficiency with respect to the union of investors' private information sets can be loosely described as a phenomenon in which the consensus belief implied by this geometric average is at least as correct as any assessment that contributes to the average. But there is no intrinsic reason why this should be the case: if two bettors have an opinion as to which horse will win a race, and one bettor's opinion is vastly superior to the other's (because of the time and expense invested in formulating that opinion), there may be no averaging of their opinions that will be superior to that of the more sagacious bettor.

This is the problem. If one investor is motivated to invest in the acquisition of information, it is within reason that he could achieve an assessment that is more accurate than that implied by the consensus belief, despite the fact that his assessment is among those that are impounded in price. To achieve a better understanding of this problem, it can be modeled along the following lines. Suppose that the mental process of determining an assessment of the return of a security is thought of as analogous to the statistical process of determining an estimate of some unknown parameter. That is, suppose that at the end of a future period, nature will draw from an urn a numbered ball that represents the return on the security. For the benefit of this discussion, it will be assumed that numbers on the balls are normally distributed in the urn with known variance and unknown mean. Before trading occurs, however, investors can draw sample observations from this urn; the observations are private information sets, and **A** can represent the union of these sets. With his private observations, each investor forms an estimate of the unknown mean which, in turn, determines his individual assessment. Then the consensus belief is at least as accurate as any single investor's assessment whenever the precision (where precision is defined to be the inverse of the variance) of the estimate of the return implicit in the consensus belief is greater than or equal to the precision in any single investor's estimate. This, in turn, implies that the market is efficient with respect to **A**, the union of investors' information sets.

The subsequent analysis considers the behavior of investors with different tolerances for risk in this situation. What is demonstrated is that whenever all investors acquire what each perceives to be an optimal quantum of information (in the form of sample observations), the degree of precision implicit in the consensus belief is indeed no less than the degree of precision in a single investor's estimate. This result implies that the Fama-Rubinstein insight as to what constitutes information efficiency is robust in that it evolves naturally from the behavior of investors operating in their own self-interests. The fact that it is robust suggests that it may provide a more viable and interesting characterization than competing theories.[4] The following analysis formally demonstrates this result.

---

without controversy that possession of more information signifies being better informed. The only way investors interact in a market is through prices; i.e., an equilibrium price is established for a security on the basis of investors' expectations, and at this price investors exchange holdings of the security. Therefore, when the price the third investor encounters is the same price that would prevail if each and every investor knew **A**, and in our previous example this was the case, there is no way he (the third investor) can earn an excess return (as I have narrowly defined the term) on the basis of **A**, despite the fact that it represents information superior to that available to the two other investors.

[4] The chief competing theory is one of "rational expectations." A recent survey article by Sanford Grossman provides a comprehensive guide to this work. The theory of rational expectations provides a particularly elegant argument for what constitutes market information efficiency. However it is also fragile in that it relies on a number of assumptions, such as the absence of "noise" (i.e., indiscernable randomness), which may not be met in a situation embodying the full complexity of a real world setting. Furthermore, it is a

I. A Description of the Market

To initiate the analysis, consider an investor i who trades in a market which offers the opportunity to purchase an unlimited amount of a risk-free security denoted by S_F, or a risky security denoted by S_R, whose total supply is S_R^T. At the end of some future period, investment in the risk-free security yields a fixed return of F units of wealth for each unit of S_F purchased, and investment in the risky security yields an uncertain return of R units of wealth for each unit of S_R purchased. At the beginning of a trading period, investor i will choose an amount S_F^i and S_R^i of the risk-free and risky security, respectively, which maximizes his expected utility subject to his initial wealth, which is denoted by W^i. That is, investor i chooses an amount S_F^i and S_R^i of the risk-free and risky security, respectively, to maximize his utility when S_F and S_R yield their respective returns.

Before trading, an investor also has the opportunity to acquire information. The information acquisition process will be characterized as a sampling process from a normal probability distribution.[5] That is, suppose that R is assumed to be some real number which is initially known to all investors to be normally distributed with mean $\bar{m}$ and precision $\bar{h}$ (where precision is defined to be the inverse of the variance $\bar{h}^{-1}$).[6] The acquisition of additional information is characterized as a sample of a predetermined size n consisting of the observation of random variables $X_1, \dots, X_n$, about which it is known initially that each is independently and identically distributed normally with mean R and precision h. Then it is well known (see Morris DeGroot, p. 167) that conditional on the observed sample values $x_1, \dots, x_n$, an investor regards the distribution of R to be normal with mean

$$\hat{m} = \frac{\bar{h}\bar{m} + nhm}{\hat{h}}$$

and precision $\hat{h} = \bar{h} + nh$

where $m = \sum_{k=1}^{n} \frac{x_k}{n}$

is the observed mean of the sample. It will also be assumed that each investor's observations are distributed independently of the observations of any other investor.

The sample mean itself, however, is not known until the information has been processed; prior to its revelation, it is a random variable. Its unconditional distribution can be determined as follows. Conditional upon knowledge of R, m is normally distributed with mean R and precision nh, because it is the mean of n sample observations, each of which is normally distributed about R with mean h. An investor's initial or prior belief about R is that it is normally distributed with mean $\bar{m}$ and precision $\bar{h}$. Thus, it can be shown that the unconditional distribution of m is a normal probability distribution with mean $\bar{m}$ and precision $nh\bar{h}/\hat{h}$.[7]

It will also be assumed that investor i has a (negative) exponential utility function of

---

notion of efficiency which is stronger than that suggested by Fama (1976), or implied by the empirical evidence concerning the existence of efficiency in the semistrong form.

[5] This characterization of information acquisition is suggested by Robert Wilson, p. 191.

[6] It should be emphasized that while prior assessments are identical, it is assumed that each investor arrives at the assessment $\bar{m}$ independent of other investors. This allows these prior assessments to be treated as though they were independent sample observations about R, each with precision $\bar{h}$.

[7] This result can be explained, intuitively, by noting that $\bar{m}$ is simply the mean of the prior distribution of R, and $nh\bar{h}/\hat{h}$ is equal to

$$\frac{nh\bar{h}}{\hat{h}} = \left[\frac{\bar{h}+nh}{nh\bar{h}}\right]^{-1} = \left[\frac{1}{nh} + \frac{1}{\bar{h}}\right]^{-1}$$

That is, the precision $nh\bar{h}/\hat{h}$ is simply the inverse of the sum of the variance $(nh)^{-1}$ of the conditional distribution of the sample mean for any given value of R, and the variance $\bar{h}^{-1}$ of the prior distribution of R. In effect, m is normally distributed with the mean investors attribute to R before they process the information, $\bar{m}$, and precision which, in a broad sense, represents the precision of their prior beliefs plus the precision of the information itself.

the form

$$U_i(w) = -r_i e^{-(1/r_i)w}$$

for wealth w, where $r_i > 0$. The coefficient $1/r_i$ is investor i's coefficient of risk aversion; its inverse r_i is referred to as his coefficient of risk tolerance. The purpose in assuming utility functions of this form is that it permits a consensus belief to be explicitly determined.[8] The extent to which the results of the analysis depend upon this assumption is not entirely clear. However, as the discussion unfolds, it becomes clear that the salient feature of the analysis is the relationship between degrees of risk tolerance and information acquisition activities; therefore, it is reasonable to conjecture that the form of the utility function assumes no more than a secondary role, provided that it evidences global risk aversion.[9]

When investor i enters the market at the beginning of the period, he encounters the following problem. If the cost associated with making n sample observations is represented by a function $C(n)$ (which is assumed to be a twice differentiable function of n), investor i must choose

1) an optimal number of sample observations, and

2) based upon the information in those observations, as well as his prior information, an optimal amount S^i_R and S^i_F of risky and risk-free securities, respectively, so as to maximize his future period expected utility; and, of course, his choices must be constrained by his initial wealth.

Suppose that a unit of the risk-free and risky security sell in the market for 1 and P units of wealth, respectively (i.e., the price of a risk-free security is a numeraire). Investor i's choice problem can be characterized mathematically as choosing those elements n, S^i_F, S^i_R that

$$(1) \quad \max_{\{n, S_F, S_R\}} \int_m \int_R -r_i e^{-1/r_i\{FS_F+RS_R\}} \cdot dN(\hat{m}, \hat{h})\, dN(\bar{m}, nh\bar{h}/\hat{h})$$

subject to $W^i = S_F + PS_R + C(n)$, where $dN(\mu, \rho)$ denotes a normal density function with mean μ and precision ρ.

For example, consider optimal amounts of S^i_F and S^i_R conditional upon the choice of n sample observations. Conditional upon n, (1) reduces to solving the Lagrangian equation

$$L(S_F, S_R, \lambda) = -\int_R r_i e^{-1/r_i\{FS_F+RS_R\}} dN(\hat{m}, \hat{h}) + \lambda\{W^i - (S_F + PS_R + C(n))\}$$

Differentiating L with respect to S_R yields

$$(2) \quad \frac{\partial L}{\partial S_R} = \int R e^{-1/r_i\{FS_F+RS_R\}} dN(\hat{m}, \hat{h}) - \lambda P$$

Differentiating L with respect to S_F yields

$$(3) \quad \frac{\partial L}{\partial S_F} = \int F e^{-1/r_i\{FS_F+RS_R\}} dN(\hat{m}, \hat{h}) - \lambda$$

Necessary first-order conditions require $\partial L/\partial S_R = \partial L/\partial S_F = 0$ in (2) and (3). Thus the two equations can be combined to yield

$$(4) \quad P = \frac{\int R e^{-1/r_i\{FS_F+RS_R\}} dN(\hat{m}, \hat{h})}{\int F e^{-1/r_i\{FS_F+RS_R\}} dN(\hat{m}, \hat{h})}$$

Recognizing that for any parameter α,

$$e^{-\alpha R} dN(\mu, v) = e^{-\alpha\mu + \alpha^2/2v} dN\left(\mu - \frac{\alpha}{v}, v\right)$$

and using a common property of the moment generating function of a normal distribution, it can be shown that (4) reduces to

---

[8] The *existence* of a consensus belief, however, can be established even in the absence of an explicit determination. See my 1979b paper.

[9] Note that the assumption of constant risk tolerance implies that an investor's tolerance for risk is independent of his wealth. However, if one postulates that risk tolerance increases as wealth increases, investors with large constant risk tolerance can be thought to assume the role of investors with large wealth.

(5)
$$P=\frac{\int R\,dN\left(\hat{m}-\frac{S_R}{r_i\hat{h}},\hat{h}\right)}{F\int dN\left(\hat{m}-\frac{S_R}{r_i\hat{h}},\hat{h}\right)}=\frac{\hat{m}-\left(S_R/r_i\hat{h}\right)}{F}$$

Rearranging terms yields the amount of the risky security investor i demands,

$$S_R^i=r_i\hat{h}\{\hat{m}-FP\} \tag{6}$$

Suppose it is assumed, however, that investor i is representative of all other investors in the market; specifically, each investor j, say, has a utility for wealth characterized by a constant tolerance for risk r_j and an assessment of the distribution of $\tilde{R}$ which is normal with mean $\hat{m}_j$ and precision $\hat{h}_j=\bar{h}+n_jh$ (where n_j is investor j's optimal number of sample observations). Then summing (6) over the demand of all investors yields

$$S_R^T=\sum_{i=1}^{N} r_i\hat{h}_i\hat{m}_i-v_0FP \tag{7}$$

where N is the total number of investors,

$$v_0\equiv\sum_{i=1}^{N} r_i\hat{h}_i$$

is the sum of the precision in each investor's assessment of the unknown mean weighted by his risk tolerance, and

$$S_R^T\equiv\sum_{i=1}^{N} S_R^i$$

represents the total supply of risky securities. Finally (7) implies

$$P=\frac{1}{F}\left\{\sum_{i=1}^{N}\left(\frac{r_i\hat{h}_i}{v_0}\right)\hat{m}_i-\frac{S_R^T}{v_0}\right\} \tag{8}$$

The expression derived in (8) is the market-clearing or equilibrium price of the risky security S_R, assuming that the price of the risk-free security is set at one.

II. Consensus Beliefs and Market Information Efficiency

To provide some intuition, the expression for P in (8) can be interpreted as the ratio of

1) the sum of each investor's assessment or estimate of the unknown mean of the return for R, $\hat{m}_i$, weighted by the coefficient $r_i\hat{h}_i/v_0$, and adjusted by a factor of supply represented by $-S_R^T/v_0$, to

2) the known fixed return of the risk-free security F.

Using the definition of a consensus belief (see Rubinstein or my 1979b paper), it is a simple exercise to show that the expression

$$\sum_{i=1}^{N}\left(\frac{r_i\hat{h}_i}{v_0}\right)\hat{m}_i$$

is a consensus belief. This is because this expression is precisely that belief, or estimate, of the unknown mean of R, which, if held by all investors in some otherwise identical market, would effect a market-clearing price identical to the one derived in (8).[10]

Consider the variance of the consensus belief. Assuming that each sample observation made by an investor is independent of all other observations made by all other investors, as well as those other observations made by himself, the variance of the consensus belief is

(9)
$$\begin{aligned}\mathrm{Var}\left[\sum_{i=1}^{N}\left(\frac{r_i\hat{h}_i}{v_0}\right)\hat{m}_i\right]&=\sum_{i=1}^{N}\left(\frac{r_i\hat{h}_i}{v_0}\right)^2\mathrm{Var}[\hat{m}_i]\\&=\sum_{i=1}^{N}\left(\frac{r_i\hat{h}_i}{v_0}\right)^2\hat{h}_i^{-1}\\&=\sum_{i=1}^{N}\left(\frac{r_i}{v_0}\right)^2\hat{h}_i\end{aligned}$$

---

[10] Specifically, it can be shown that if each investor regards the distribution of R to be normal with mean $\Sigma(r_i\hat{h}_i/v_0)\hat{m}_i$ and precision $v_0/(\Sigma r_i)$, a price identical to the one in equation (8) obtains.

remembering that the $r_i\hat{h}_i$ are all constant.[11]

As discussed heuristically in the introduction, the notion of market information efficiency that underlies the concept of a consensus belief is that the variance in the consensus belief assessment of uncertainty (in this case the unknown mean of R) is less than or equal to the variance in each investor's personal assessment. (This, of course, implies that the precision in the consensus belief is at least as great as the precision in each investor's assessment.) Since the variance of the consensus belief assessment is given by (9), and the variance of each investor's assessment is given by $\hat{h}_j^{-1}$, this can be expressed formally as the requirement that for each investor $j, j=1,\ldots,N$,

$$(10) \qquad \sum_{i=1}^{N}\left(\frac{r_i}{v_0}\right)^2\hat{h}_i \leqslant \hat{h}_j^{-1}$$

The interesting question is under what circumstances the expression in (10) would hold. Casual observation of (10) suggests that if investors with large risk tolerance have little precision in their assessments, and vice versa, then the variance in the consensus belief will be large, because in effect poor assessments are given greater weight in determining the consensus belief than good assessments. Therefore, for (10) to hold, it must be that investors with greater risk tolerance somehow contribute better assessments to determining the market-clearing price than investors with lower risk tolerance. The next section considers how investors choose an optimal degree of precision on the basis of their risk tolerance, and suggests a circumstance in which the appropriate behavior results.

III. A Sufficient Condition to Ensure Information Efficiency

The characterization of market efficiency implied by the consensus belief interpretation requires that equation (10) hold. The validity of equation (10) is demonstrated in a number of steps. First, a function $H(r)$ is introduced to represent an investor's optimal degree of precision as a function of his tolerance for risk through the relationship

$$H(r)=\hat{h}=\bar{h}+n_0(r)h$$

where $n_0(r)$ represents the optimal number of sample observations an investor requires as a function of his risk tolerance. Second, in this section, it is demonstrated that a sufficient condition that equation (10) hold is that $H(r)$ is a nondecreasing function of r; that is, each investor's optimal degree of precision is a nondecreasing function of his risk tolerance. Finally, the fact that $H(r)$ is indeed a nondecreasing function of r is demonstrated in the next section.

If it is assumed that $H(r)$ is nondecreasing, proving the validity of equation (10) is straightforward. To begin, consider the investor whose risk tolerance is no less than the risk tolerance of any investor in the market; this investor is referred to as investor k. That is, $r_k > r_i$ for all i. This implies that for all i,

$$r_iH(r_k) \leqslant r_kH(r_k) \leqslant \sum_{i=1}^{N} r_iH(r_i) \equiv v_0$$

which in turn implies that for all i,

$$(11) \qquad H(r_k) \leqslant \frac{v_0}{r_i}$$

However, for equation (10) to be valid for investor k, it must be that

$$\sum_{i=1}^{N}\frac{r_i^2H(r_i)}{v_0^2} \leqslant \{H(r_k)\}^{-1}$$

or

$$(12) \qquad \sum_{i=1}^{N}\frac{r_i^2H(r_i)H(r_k)}{v_0^2} \leqslant 1$$

Substituting the inequality in (11) into the left-hand side of (12) yields

---

[11] This also requires that each investor's prior assessment of the mean of R, $\bar{m}$ is arrived at independent of other investors. See fn. 6.

$$\sum_{i=1}^{N} \frac{r_i^2 H(r_i) H(r_k)}{v_0^2} \leqslant \sum_{i=1}^{N} \frac{r_i^2 H(r_i)}{v_0^2} \cdot \frac{v_0}{r_i}$$

$$= \sum_{i=1}^{N} \frac{r_i H(r_i)}{v_0} = 1$$

which implies (12). Thus, the degree of precision in investor k's assessment is no greater than the degree of precision in the consensus belief assessment.

But note that if $H(r)$ is a nondecreasing function of r, then it must be that for all i, $H(r_k) \geqslant H(r_i)$. That is, investor k attains a degree of precision that is no less than the degree attained by anyone else in the market. Consequently, no one else in the market has a degree of precision which is greater than the degree of precision in the consensus belief assessment, which ensures equation (10). In brief, market information efficiency holds in the sense of (10) whenever each investor attains as an optimal degree of precision a level which is a nondecreasing function of his risk tolerance.

It remains to show that $H(r)$ is a nondecreasing function. However, even in the absence of the analysis performed below, one would intuitively expect this to be the case. This is because investors with greater risk tolerance invest a greater proportion of their wealth in the risky security relative to investors with the same wealth and the same values for mean and precision, but with lower tolerances for risk. For example, this relationship can be seen from the expression for investor demand derived in (6):

$$S_R^i = r_i \hat{h}_i \{\hat{m}_i - FP\}$$

Therefore to safeguard their increased amount of investment in the risky security, one might reasonably expect risk tolerant investors to acquire more information. This is shown in the next section.

IV. Optimal Sampling Sizes

In this section it is formally demonstrated that $n_0(r)$ is a nondecreasing function of r; that is, an investor's optimal number of sample observations is a nondecreasing function of his risk tolerance. Since the optimal degree of precision is a linear function of the optimal sample size, this implies that the optimal degree of precision is a nondecreasing function as well.

To show that $n_0(r)$ is a nondecreasing function of r, an expression for equation (1) is needed. Recall that since it was shown that $S_R^i = r_i \hat{h}\{\hat{m} - FP\}$ and $S_F^i = W^i - C(n) - PS_R^i$ are optimal amounts of the risky and risk-free security, respectively, (1) can be expressed as

$$\max_{\{n\}} G(n, r)$$

$$\text{subject to} \quad W = S_F + PS_R + C(n)$$

where $G(n, r)$

$$= \int_m \int_R -re^{-F/r\{W - C(n))\} - \hat{h}\{R - FP\}\{\hat{m} - FP\}}$$

$$\times dN(\hat{m}, \hat{h})\, dN(\bar{m}, nh\bar{h}/\hat{h})$$

and the i sub- and superscripts used to identify investor i are also dropped to ease the notational burden. That is, $G(n, r)$ is the expected utility an investor with risk tolerance r achieves when he chooses to make n sample observations.

Using the properties of a normal moment generating function discussed above, integration with respect to R yields

(13)

$$G(n, r) = \int_m -re^{-F/r\{W - C(n)\} - \frac{1}{2}\hat{h}\{\hat{m} - FP\}^2}$$

$$\times dN(\bar{m}, nh\bar{h}/\hat{h})$$

As a means of integrating the expression in (13), let the variable y be introduced, where y is defined as

$$y = (nh\bar{h}/\hat{h})^{1/2}\{m - \bar{m}\}$$

which implies $dy = (nh\bar{h}/\hat{h})^{1/2}\, dm$

Also note that a priori an investor does not know what value P will assume. It will be assumed here that each investor will behave *as if* he believes that price will simply reflect the ratio of what all investors know initially about the expected return for R, i.e., $\bar{m}$, to the risk-free return F; that is, all investors will behave as if they believe that $P=\bar{m}/F$. Stated differently, investors act as if prices reflect only what is already well known, and not (among other things) their own additional information acquisition activities, or the information acquisition activities of others.[12] This assumption is consistent with the notion that as price takers in a traditional market setting, investors behave as if they perceive no relationship between their own activities and price. Analytically, this assumption implies (with some algebraic manipulation) that

$$\hat{h}\{\hat{m}-FP\}^2=\{(nh)^2/\hat{h}\}\{m-\bar{m}\}^2$$

$$=\{nh/\bar{h}\}y^2$$

where the variable y is used as a substitution.

It can be shown that a further substitution of y in place of the variable m transforms (13) into

$$G(n,r)$$

$$=\int_y -re^{-F/r\{W-C(n)\}-1/2\{1+(nh/\bar{h})\}y^2}dy$$

which, when integrated using properties of a normal probability distribution, yields

$$G(n,r)=-r(\bar{h}/\hat{h})^{1/2}e^{-F/r\{W-C(n)\}}$$

With an expression for $G(n,r)$ now available, define a function $n_0(r)$ which represents the value of n which maximizes $G(n,r)$ for each value of r. The function $n_0(r)$ represents the optimal number of sample observations an investor should acquire when his risk tolerance is r. It is necessary to show that $n_0(r)$ is a nondecreasing function of r. This is done using the above expression for $G(n,r)$.

Allowing the derivative with respect to n to be denoted by an asterisk, observe that

$$G^*(n,r)=\left\{\frac{F}{r}C^*(n)-\frac{1}{2}h(nh+\bar{h})^{-1}\right\}$$

$$\times G(n,r)$$

since

$$\frac{\partial}{\partial n}(\bar{h}/\hat{h})^{1/2}=-\frac{1}{2}(\bar{h})^{1/2}h(nh+\bar{h})^{-3/2}$$

$$=-\frac{1}{2}h(nh+\bar{h})^{-1}(\bar{h}/\hat{h})^{1/2}$$

Assuming that $C(n)$ is a strictly increasing function of n, a necessary first-order condition for maximization requires that $n_0(r)$ satisfy

$$(14)\quad C^*(n_0(r))=\frac{rh}{2F\{n_0(r)h+\bar{h}\}}$$

since in this case $G^*=0$. Furthermore,

$$(15)\quad G^{**}(n,r)$$

$$=\left\{\frac{F}{r}C^{**}(n)+\frac{1}{2}h^2(nh+\bar{h})^{-2}\right\}G(n,r)$$

$$+\left\{\frac{F}{r}C^*(n)-\frac{1}{2}h(nh+\bar{h})^{-1}\right\}^2G(n,r)$$

Therefore, a sufficient second-order condition for maximization is that for all n,

$$(16)\quad \frac{F}{r}C^{**}(n)+\frac{1}{2}h^2(nh+\bar{h})^{-2}>0$$

since in this case G is a concave function. (Concavity follows from the fact that $G(n,r)$ is negative for all n and r, and the expressions in (15) which precede those of $G(n,r)$ are both positive; thus G^{**} is negative whenever (16) holds.) Of course, maximization also requires that there is sufficient wealth for an investor to achieve this optimum, i.e.,

---

[12] This assumption is incompatible with theories, such as that of rational expectations, which rely substantially upon investors' ability to decode prices, or infer the information acquired by others through prices, over time (see Grossman). In effect, it suggests that investors evidence somewhat myopic behavior. Its advantage is that it may describe what occurs in an actual market setting.

$$PS_R + C(n_0(r)) \leqslant W$$

But this will also be assumed since we are specifically concerned with the circumstance in which each investor achieves his optimal sample size.[13]

As an aside, note that while (16) does not require that the cost function is convex, for all practical purposes it does since otherwise a counterexample to the claim that G is concave can likely be constructed by an appropriate choice of parameters. This limitation notwithstanding, consider the objective of a cost function more generally. For equation (10) to hold, the cost function must prohibit a low-risk-tolerant investor from attaining a greater degree of precision than a high-risk-tolerant investor, or else the (geometric) average of their precisions may no longer exceed the precision of the investor with greatest precision. Increasing marginal cost of increased sampling is one circumstance which achieves this.

Finally, we demonstrate that the behavior of an optimal number of sample observations, as expressed in equation (14), is sufficient to imply that $n_0(r)$ is a nondecreasing function of r. Assuming that n_0 is differentiable with respect to r, differentiating both sides of (14) yields (where differentiation with respect to r is denoted by a prime):

$$C^{**}(n_0)\cdot n_0' = \frac{h}{2F}\left\{\frac{1}{(n_0h+\bar{h})} - \frac{rhn_0'}{(n_0h+\bar{h})^2}\right\}$$

or, rearranging terms,

$$(17)\quad n_0' = \frac{\dfrac{h}{2F(n_0h+\bar{h})}}{C^{**}(n_0) + \dfrac{rh^2}{2F(n_0h+\bar{h})^2}}$$

But the numerator of (17) is clearly nonnegative because all its terms are; the denominator is positive under the assumption ensuring concavity of G expressed in (16). Therefore, n_0' is a positive function of r, which implies that n_0 is (at least) a nondecreasing function of r.

The fact that $n_0(r)$ is nondecreasing immediately implies that $H(r)$ is also nondecreasing through the relationship

$$H'(r) = n_0'(r)h$$

In summary, if each investor can attain his optimal precision by acquiring his optimal number of sample observations, the degree of precision will be a nondecreasing function of his risk tolerance. As discussed previously, this behavior is sufficient to ensure that a market is informationally efficient in that equation (10) holds.

V. Conclusion

The interesting elements of any characterization of information efficiency are those conditions that are sufficient to induce it. This analysis attempts to point out that although there is no intrinsic reason why the geometric averaging performed by prices should yield an assessment (i.e., the consensus belief) which is no less precise than the individual assessments which contribute to the average, it holds whenever all investors acquire what each perceives to be an optimal quantum of information in the presence of some common cost function. For example, suppose that publicly available information is thought of as an urn from which investors make sample observations at some cost (in wealth, time, or effort). Then, within the context of the notion of market efficiency discussed here, a market efficient with respect to all publicly available information would naturally result; sufficient conditions would require no more than that investors act in their own self-interest. But, investors acting in their own self-interest is an assumption common to most discussions of competitive markets. Therefore, this interpretation of market information efficiency may imply a phenomenon that arises naturally from the market process.

---

[13] Note that given the nature of each investor's choice problem, the nonnegative amount $W-PS_R-C(n)$ which remains after n and S_R are selected is simply that amount allocated to investment in the risk-free security.

Of course, there are a number of assumptions made in this analysis that limit generalization. Most significantly, it is assumed that investors face a common cost function which, although not necessarily convex, may approximate that requirement. A further restriction, but less obvious in its impact, is the assumption that all investors have constant risk tolerance. Thus, it is easy to imagine situations in which efficiency within the context of the consensus belief interpretation will not obtain. However, situations of this nature may be useful for explaining anomalies in the literature of efficient markets. The point of this paper is to demonstrate that the consensus belief interpretation of market information efficiency is both viable and robust.

APPENDIX

S_F = the risk-free security

S_R = the risky security

S_R^T = the total supply of the risky security

F = the risk-free rate of return

R = the (uncertain) risky rate of return

i = the ith investor

N = the total number of investors

W^i = the endowed wealth of investor i

S_F^i = the amount of investment in the risk-free security which is optimal for investor i

S_R^i = the amount of investment in the risky security which is optimal for investor i

$\bar{m}$ = the mean of the distribution of R prior to acquiring additional information

$\bar{h}$ = the precision of the distribution of R prior to acquiring additional information

m = the mean of a sample of observations from a process distributed normally with mean R and precision h

nh = the precision of a sample of n observations from a process distributed normally with mean R and precision h

$\hat{m}$ = the mean of the distribution of R posterior to acquiring additional information

$\hat{h}$ = the precision of the distribution of R posterior to acquiring additional information

r_i = the risk tolerance of investor i

$C(n)$ = the cost of sampling n observations

$dN(\mu, \rho)$ = a normal density function with mean μ and precision ρ

v_0 = the value of the sum over all investors of the product $r_i \hat{h}_i$

$n_0(r)$ = the optimal number of sample observations acquired by an investor with risk tolerance r

$H(r)$ = the optimal degree of precision attained by an investor with risk tolerance r, i.e., $H(r) = \bar{h} + n_0(r)h$

$G(n, r)$ = the expected utility an investor with risk tolerance r achieves when he chooses n sample observations.

REFERENCES

W. H. Beaver, "Market Efficiency," work. paper, Stanford Univ., Apr. 1978.

A. Beja, "The Limited Information Efficiency of Market Processes," finance work. papers, Univ. California, Feb. 1976.

Morris H. De Groot, *Optimal Statistical Decisions*, New York, 1970.

Eugene F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, May 1970, *25*, 383–417.

______, *Foundations of Finance*, New York 1976.

S. Figlewski, "Market 'Efficiency' in a Market with Heterogeneous Information," *J. Polit. Econ.*, Aug. 1978, *86*, 581–97.

S. J. Grossman, "Rational Expectations and the Allocation of Resources under Asymmetric Information: A Survey," work. paper no. 79–18, Univ. Pennsylvania 1979.

______ **and J. E. Stiglitz,** "Information and Competitive Price Systems," *Amer. Econ. Rev. Proc.*, May 1976, *66*, 246–53.

M. Rubinstein, "Securities Market Efficiency in an Arrow-Debreu Economy," *Amer.*

Econ. Rev., Dec. 1975, *65*, 812–24.

R. E. Verrecchia, (1979a) "On the Theory of Market Information Efficiency," *J. Accounting Econ.*, Mar. 1979, *1*, 77–90.

———, (1979b) "A Proof of the Existence of 'Consensus Beliefs'," *J. Finance*, Sept. 1979, *34*, 957–63.

R. B. Wilson, "Informational Economies of Scale," *Bell J. Econ.*, Spring 1975, *6*, 184–95.

Unemployment Spells and Unemployment Experience

By George A. Akerlof and Brian G. M. Main*

According to a modern view of the labor market, as reflected most prominently in papers by Martin Feldstein, Robert Hall, and George Perry, among others, unemployment in the United States is on average of quite short duration and, as a result, should be modeled not in terms of stocks of persons who are unemployed for fairly long periods of time, but rather in terms of flows of persons whose spells of unemployment are quite short.

An important statistic which supports this view of unemployment is the large number of completed spells of unemployment in any one year. For example, in the year 1969, which is in no way atypical in this respect, 93,640,000 persons[1] were in the labor force some fraction of the time, and there were a reported 32,100,000 spells of unemployment with an average duration of only 4.6 weeks.[2] This paper examines ways in which such statistics on the large number of spells and the correspondingly short durations are misleading in terms of the welfare implications of unemployment.

Note that the average duration statistic just cited concerned *completed* spells of unemployment. It is not the same as the usually quoted official government statistic, the average duration of spells currently in progress (or *interrupted* spells). The difference between these two statistics is exactly analogous to the difference between mean life span (which is the average length of a completed life) and the mean age of the population (which is the average length of lives currently in progress). This analogy between mean life span and mean age suggests that the average duration of spells in progress should underestimate the average length of completed spells because the duration of each spell in progress is less than its completed length. However, there is also a bias in the other direction. In any sample of the population, a spell in progress has a probability of being sampled which is proportional to its length, and thus the sampled average length of spells in progress may overstate the length of completed spells since the longer spells are proportionately oversampled.[3] The paper by Hyman Kaitz, which for the first time estimated the average length of completed spells for the United States, has been quite influential in supporting the new view of unemployment. His paper shows that the average length of a completed spell from 1948 to 1969 was only 40–62 percent of the average length of spells in progress. Table 1 brings Kaitz' estimates on the average lengths of completed spells up to date using the method proposed by Steven Salant.[4] It also compares the estimated average completed spell lengths for these years to the official statistics on average in-progress (or interrupted) spell lengths.

This paper concerns the interpretation of a spell of unemployment, and offers new information on the length of spells of unemployment separately for people who experi-

---

*University of California-Berkeley. We would like to thank James Berry, Norman Bowers, Harvey Hamel, Steve Salant, and Janet Yellen for valuable comments and assistance. We also gratefully acknowledge the generous financial support of the Board of Governors of the Federal Reserve System and the U.S. Department of Labor under Small-Grant Research Project No. 91-06-78-27, administered by the Institute of Business and Economic Research of the University of California-Berkeley.

[1] See Table C-5, U.S. Bureau of Statistics (1970).

[2] See Kaitz, Table 2, p. 12.

[3] For a good discussion of these issues, see Salant.

[4] According to Salant, if the density of interrupted spell lengths T is $g(T)=(r-1)a^{r-1}(a+T)^{-r}$, the density of completed spell lengths S is $f(S)=ra^{r}(a+S)^{-(r+1)}$. Data on interrupted spells were used for maximum likelihood estimates of a and r. The function $f(S)$ with these estimated values yields estimated densities of completed spell lengths.

TABLE 1—DISTRIBUTION OF COMPLETED UNEMPLOYMENT SPELLS BY DURATION[a]

Year	Total	Number of Weeks					Average Duration[b]	Interrupted Spell Average Duration[b]	Unemployment Rate
		<5	5–10	11–14	15–26	>26			
1948	25,580	18,880	4,190	900	1,140	470	4.6	8.6	3.8
1949	30,450	20,310	5,150	1,630	2,320	1,040	6.2	10.0	5.9
1950	25,510	16,870	4,150	1,270	2,050	1,170	6.7	12.1	5.3
1951	26,010	20,700	3,000	780	1,070	460	4.1	9.7	3.3
1952	24,540	19,150	3,510	680	850	350	4.0	8.4	3.0
1953	25,580	20,470	3,370	630	790	320	3.7	8.0	2.9
1954	28,600	18,720	5,300	1,320	2,020	1,240	6.4	11.8	5.5
1955	27,300	19,720	4,360	860	1,360	1,030	5.4	13.0	4.4
1956	28,680	21,010	4,480	960	1,420	810	5.0	11.3	4.1
1957	26,190	18,150	4,420	1,130	1,650	840	5.7	10.5	4.3
1958	32,350	21,890	4,160	1,470	2,650	2,180	7.4	13.9	6.8
1959	31,220	22,190	4,200	1,190	2,180	1,460	6.2	14.4	5.5
1960	33,590	24,190	4,280	1,380	2,320	1,420	6.0	12.8	5.5
1961	34,230	23,060	5,220	1,330	2,450	2,170	7.2	15.6	6.7
1962	32,900	22,980	5,230	1,150	1,930	1,610	6.2	14.7	5.5
1963	33,210	22,590	5,500	1,320	2,220	1,580	6.4	14.0	5.7
1964	33,860	24,130	5,090	1,200	2,010	1,430	5.8	13.3	5.2
1965	33,670	25,000	4,610	1,120	1,810	1,130	5.2	11.8	4.5
1966	35,900	28,410	4,420	900	1,370	800	4.2	10.4	3.8
1967	31,110	22,090	5,680	1,160	1,500	680	5.0	8.8	3.8
1968	32,340	23,980	5,390	1,020	1,320	630	4.5	8.4	3.6
1969	32,100	23,670	5,360	1,110	1,390	570	4.6	7.9	3.5
1970	39,120	24,550	9,260	2,240	2,220	850	5.4	8.7	4.9
1971	38,010	21,570	9,460	2,600	2,910	1,470	6.8	11.3	5.9
1972	40,750	25,230	9,260	2,350	2,550	1,360	6.2	12.0	5.6
1973	31,980	17,230	8,420	2,410	2,710	1,210	7.0	10.0	4.9
1974	47,150	29,630	10,940	2,670	2,730	1,180	5.6	9.7	5.6
1975	44,570	21,240	11,730	3,770	4,810	3,020	9.1	14.1	8.5
1976	47,600	26,610	11,370	3,210	3,850	2,560	8.0	15.8	7.7
1977	49,540	26,100	13,200	3,870	4,390	1,980	7.2	14.3	7.0
1978	50,080	30,050	11,770	3,050	3,320	1,690	6.3	13.1	6.0

[a] Shown in thousands.
[b] Shown in weeks.

ence single spells, two spells, or more than two spells of unemployment in a single year. The key to our interpretation is that the incidence of unemployment, or indeed of the labor market conditions which cause unemployment, on individuals who suffer from unemployment cannot simply be associated with the duration of completed spells of unemployment for at least two reasons. First, there are many persons who, even over a short period (a calendar year, for example) suffer from *multiple* spells of unemployment. For them the duration of a *completed* spell understates their unemployment experience.[5] Secondly, the impact of bad labor market conditions on individuals may not end with a spell of unemployment, if at the end of that spell the worker has adjusted to these conditions either by taking temporary work or by dropping out of the labor force.[6]

I. Single Spells and Multiple Spells Contrasted with the Completed Spell

The durations of a completed spell of unemployment are meaningful in measuring the incidence of unemployment for persons with only one spell of unemployment over the period of observation (for example, over a calendar year). However, it fails to measure the incidence of unemployment for persons with repeat spells. Such repeat spells, of course, could be ignored if relatively few persons suffered from them or if they were relatively small in number. The contrary, however, appears to be the case. In 1976, for example, according to the annual *Work Experience Survey* of the Bureau of Labor Statistics, 34 percent of all persons with some unemployment had more than one spell and more than half of all spells were suffered by such persons.

---

[5] Kaitz mentions the importance of repeat spells, especially among people with short spells: "Among the 24 million spells of less than five weeks, many were completed by individuals who had more than one spell" (p. 12).

[6] For an analysis of this question, see the excellent paper by Kim Clark and Lawrence Summers.

The use of the completed spell of unemployment in support of the "new view" theory of the labor market is most convincing if all spells corresponded to a stereotypical spell, in which the person out of work is continuously searching for a new job and the unemployment is sandwiched between jobs of long tenure. In this case, the person involved will suffer only a single spell of unemployment over a fairly long period of time, so that the period of unemployment and of undesired labor force status of the individual coincide. This coincidence is much less clear cut in the more common spells of unemployment which do not correspond to the stereotype.

Unemployment for persons with multiple spells could be loosely classified into two types. First, many persons with short job tenures are in industries whose jobs are organized to have short duration. In 1976, for example, 18.0 percent of persons with three or more spells of unemployment were in the construction industry, whereas construction workers constituted only 5.6 percent of those with work experience.[7] For such persons, spells of unemployment are inherent in the organization of their industry.

There are also industries in which jobs are typically of long tenure; these persons, on losing such a job, seek another. The Bureau of Labor Statistics records a person as unemployed if he is reported out of work but has searched for work in the last four weeks. The person is considered "out of the labor force" if he did not work and he did not search. According to the Current Population Survey *Interviewers' Reference Manual* (U.S. Census, 1976) on the employment duration question, the current period of unemployment is to begin with the current period of search.[8] Thus failure to search for a job, out of discouragement or knowledge that no suitable job is available, or because times of unemployment involve relatively little opportunity cost for a vacation or for some domestic project, breaks a person's recorded spell of unemployment, provided census interviewers successfully follow their instructions. For men, one-third, and for women, more than one-half, of all spells of unemployment end out of the labor force (see S. H. Garfinkle, p. 56, Table 4). And approximately 40 percent of all spells of unemployment begin from out of the labor force, rather than from employment (see John Bregger, Table 2).

Nor do interruptions in a spell of recorded unemployment by a spell of recorded employment necessarily signify the availability of jobs at which a worker can earn his permanent income. A person who has searched long enough to know that there are no permanent jobs commensurate with his skills may well take temporary work at considerably reduced wages or with considerably reduced nonpecuniary rewards.[9] There are a large number of jobs of short

---

[7] See U.S. Bureau of Labor Statistics (1977).

[8] The following passage is the instruction to *CPS* interviewers regarding the beginning of the period of unemployment:

C. Continuously looking: The number of weeks should represent the *present continuous* period of looking for work. Consider a period of two weeks or more during which a person was employed or ceased looking for work as breaking the continuity of the present period of seeking work. In such cases, enter the number of weeks of the *present* period of looking for work, counting from the end of the "break in continuity" through end of survey week.

Example: Arthur Hodges looks for work for four weeks, then his uncle takes him on a vacation trip for four weeks. Arthur does not look for a job again until the beginning of survey week. The entry in item 22C is "01," representing the present continuous period of looking for work. [U.S. Bureau of the Census, p. D6-40]

[9] Additionally, workers facing long periods of layoff may take relatively low paying jobs which they view as temporary.

TABLE 2—UNEMPLOYMENT EXPERIENCE: ALL PERSONS WITH WORK EXPERIENCE

Year	Average Annual Experience of Unemployment for People With[a]			Average Spell Length for People With[b]			Average Length of Completed Spell[b]
	1 Spell	2 Spells	>2 Spells[c]	1 Spell	2 Spells	>2 Spells[c]	
1965	10.1	14.8	18.1	10.1	7.4	5.2	7.3
1966	9.2	13.6	16.8	9.2	6.8	4.8	6.7
1967	9.1	13.3	16.5	9.1	6.7	4.7	6.7
1968	8.4	13.3	16.5	8.4	6.7	4.7	6.6
1969	9.0	12.4	16.5	9.0	6.2	4.7	6.6
1970	11.5	15.4	19.1	11.5	7.7	5.5	8.3
1971	13.4	17.6	20.4	13.4	8.8	5.8	9.5
1972	12.7	16.4	19.4	12.7	8.2	5.5	8.9
1973	11.0	15.2	18.3	11.0	7.6	5.2	8.0
1974	11.6	15.7	18.0	11.6	7.9	5.1	8.2
1975	15.4	19.2	19.4	15.4	9.6	5.6	10.7
1976	14.6	18.4	20.3	14.6	9.2	5.8	10.2
1977	13.6	17.1	18.8	13.6	8.6	5.4	9.5
Average	11.5	15.5	18.3	11.5	7.8	5.2	8.2

Source: *BLS Work Experience of the Population*, various years.

[a] In weeks per year.

[b] In weeks per spell.

[c] Assumes average of greater than 2 spells category is 3.5 spells.

duration. According to the most recent *Special Labor Force Report* on *Job Tenure of Workers*, in January 1973, 17.3 percent of all persons were at work on jobs which had lasted six months or less. A finer breakdown on short-term jobs was not available. The median length of time on the current job was only 3.9 years.[10]

Since the durations of completed spells are useful in calculating the incidence of unemployment for persons with only one spell of unemployment over a period of observation, we classified the unemployment experiences of persons according to the number of spells of unemployment which they experienced over the course of a calendar year. The results of this classification taken from the annual *Work Experience Survey* are reported in Table 2. They reveal an empirical regularity which has implications for persons with all types of unemployment experience. Every group, persons with single spells, double spells, and three or more spells, experienced more unemployment than indicated by the average length of a completed spell. Persons with single spells have more unemployment because multiple spells are on the average shorter than single spells; persons with more than one spell have more unemployment because of the multiplicity of their spells.

The average durations in Table 2 were calculated from data on the numbers of persons with work experience, with unemployment experience lasting 0–5 weeks, 6–10 weeks, 11–14 weeks, 15–26 weeks, and 27–52 weeks. The average duration of unemployment experience for persons in each of these intervals was taken to be its midpoint. The crudeness of this assumption should not affect the relative values of the average spell lengths of persons with one, two, or three or more spells of unemployment since the same assumption was used consistently throughout.

According to Table 2 for all the years of the annual *Work Experience Survey* from 1965 to 1977, for the population with some work over the course of the year, the average duration of unemployment spells decreases with the number of spells experienced in the calendar year. Total weeks of unemployment over the year increases, but considerably less than proportionately with the number of spells of unemployment. The average length of unemployment experienced increases only 37 percent from one to two spells, and only 18 percent from two spells to three or more spells. Correspondingly, the duration of double spells was on average 31 percent shorter than the duration of single spells, and the duration of "triple-plus" spells was on average 32 percent shorter than the duration of double spells.

---

[10] See U.S. Bureau of Labor Statistics (1975, p. A-8, Table A).

In the case of persons with single spells there is no ambiguity as to whether the person involved belongs to an industry organized in such a way that multiple spells over the course of a calendar year are automatic. The average length of time for such persons between long-term jobs is understated by the average durations of their spells. There are two reasons: the interruption of any spells of unemployment which happen to be in progress either at the beginning or end of the calendar year; and the existence of other spells not recorded due to the arbitrary division imposed by the calendar-year observation period. There is, furthermore, a systematic reason why persons with short periods between long-term jobs will have single spells of unemployment and persons with long periods between long-term jobs will have multiple spells; the interim adjustment to such a long spell between long-term jobs, either by dropping briefly out of the labor force or by taking temporary work, breaks the spell of unemployment. For these reasons it is significant that single spells are considerably longer than the average completed spell of unemployment. The average durations of single spells constitutes a conservative lower bound on the average durations of periods suffered by persons between long-term jobs.

II. Tests of Robustness of Negative Correlation between Length and Number of Spells

The previous section reported an empirical regularity regarding the negative correlation between average spell lengths and the number of spells of unemployment experienced in a calendar year. This section will discuss potential biases in this data with an attempt to estimate the order of magnitude of those biases. As it turns out, none of these biases significantly modify the conclusion that persons with single spells have significantly longer unemployment durations than persons with multiple spells.

A. *Reporting Bias*

The results of Table 2 are at least a bit questionable because of its reliance on data from the annual work experience surveys. These surveys appear to be quite consistent with the month-to-month *CPS* with respect to the reporting of the unemployment experience of prime age and of older men, but both women and teenagers report significantly fewer weeks of unemployment experience than is consistent with the monthly interviews. For example, for the years 1964 to 1971 the weeks of unemployment of women according to the *Work Experience Survey* understates the weeks of unemployment according to the *CPS* by an average of 19 percent. (See Richard Morgenstern and Nancy Barrett, p. 356.) For that reason it was felt desirable to construct a separate table for the unemployment experience of prime age men, whose unemployment is overreported by an average of only 10 percent.[11]

Unpublished data were available from the Bureau of Labor Statistics to construct Table 3 which reports estimates of unemployment duration of men 25–45 and over 45, respectively, for the years 1974–1977 for persons with some work experience. It is significant that these groups display the same negative correlation between average durations of spells of unemployment and the number of spells that occur for the population as a whole. Among older men, for example, those with two spells of unemployment have only 10 percent more unemployment than persons with one spell, and older

---

[11] For a similar point see June O'Neill.

TABLE 3—UNEMPLOYMENT EXPERIENCE: MALES AGED 45 AND OVER AND 25–44

	Average Annual Experience of Unemployment[a]			Average Spell Length[b]			Average Length of Completed Spell
Year	1 Spell	2 Spells	>2 Spells[c]	1 Spell	2 Spells	>2 Spells[c]	
			A. Males Aged 45 and Over				
1974	14.9	16.6	20.3	14.9	8.3	5.8	9.9
1975	18.9	20.6	18.8	18.9	10.3	5.4	11.4
1976	19.2	19.5	19.7	19.2	9.7	5.6	11.4
1977	17.2	20.3	19.1	17.2	10.2	5.5	11.1
Average	17.6	19.3	19.5	17.6	9.6	5.6	11.0
			B. Males Aged 25–44				
1974	12.6	16.1	18.0	12.6	8.1	5.2	8.3
1975	16.3	20.1	20.0	16.3	10.1	5.7	11.1
1976	15.6	18.2	21.2	15.6	9.1	6.1	10.3
1977	14.9	17.6	19.7	14.9	8.8	5.6	10.1
Average	14.9	18.0	19.7	14.9	9.0	5.6	10.0

Source: Unpublished data supplied by Bureau of Labor Statistics.
[a,b,c]See Table 2.

men with three or more spells have only 2 percent more, while average durations decline by 45 percent from one to two spells and 49 percent from two to three or more spells. The qualitative agreement between these groups of prime age men and the whole population indicates that the negative correlation between spell length and number of spells is not merely due to the failure of the unemployment of women and teenagers to be recalled correctly in the annual *Work Experience Survey*.

B. *Interruption Bias*

There is a second bias in the data which concerns the period of observation (i.e. one calendar year). Not all the spells of unemployment of the whole population reported in Table 2 or of the adult males reported in Table 3 were recorded in their entirety, since some spells were interrupted by the beginning of the calendar year for which the work experience was recorded and other spells were interrupted by the end of the calendar year. The introduction to this paper focused on the meaning of the length of a completed spell. To be consistent with that introduction the statistics presented should only concern *completed* spells. Direct observation of such spells, unmixed with interrupted spells, is only available from longitudinal data (see Parnes et al., 1968, 1972a, b). Four sets of such data are available, for young men, young women, and older men. Because of the underreporting biases with respect to youth and women, our attention is focused on older men. From the National Longitudinal Survey it was possible to compute the average length of completed spell for spells totally experienced in the interview year 1966–67. These spell lengths are classified in Table 4 according to the number of spells completed within the interview year. The qualitative results of Tables 2 and 3 hold up: unemployment experience increases with the number of spells but considerably less than proportionately.

It is also of interest to compare the average length of a single spell with the average length of all completed spells. This spell is 57 percent longer. Thus for a person whose unemployment experience in a given year is unambiguously measured by the length of a single spell, then this spell is on average 57 percent longer than the average length of all completed spells.

C. *Finite Period of Observation Bias*

The data presented have shown that single spells of unemployment have average dura-

TABLE 4—UNEMPLOYMENT EXPERIENCE: MALES AGED 45–59

	Average Annual Experience of Unemployment[a]			Average Spell Length[b]			Average Length of Completed Spell
Year	1 Spell	2 Spells	>2 Spells[c]	1 Spell	2 Spells	>2 Spells[c]	
1966–67	9.9	12.0	14.0	9.9	6.0	4.0	6.3

Source: Parnes' National Longitudinal Survey of Labor Force Experience.
[a,b,c]See Table 2.

TABLE 5—SIMULATED UNEMPLOYMENT EXPERIENCE OF 100,000 MALES AGED 45–59, 1966–67

Average Annual Experience of Unemployment for Those with[a]			Average Spell Length for Those with[b]			Average Length of Completed Spell
1 Spell	2 Spells	3 Spells	1 Spell	2 Spells	3 Spells	
6.1	10.8	13.1	6.1	5.4	4.4	6.0

Source: A simulation of the labor market experience of 100,000 men over one year based on the average behavior of 4,500 men from Parnes' National Longitudinal Survey of Labor Market Experience, 1966–67.
[a,b]See Table 2.

tions which are greater than the average durations of repeated spells and, consequently, of all completed spells. Moreover, the simplest null hypothesis that all persons have equal transition probabilities between different labor market states predicts, similarly, that over any finite period of observation persons with more spells of unemployment will have on average shorter spells; the shortness of a single spell increases the probability of having other spells in the remaining time in the finite period of observation. No data which is conceivably available would show directly *how much shorter* multiple spells will be, due to the finite observation period.

It was, however, possible to perform a simulation to approximate the order of magnitude of this bias. Our prior paper used the same sample of older men from the National Longitudinal Survey for the same interview year 1966–67 whose completed spells were reported in Table 4. That paper computed the average weekly probability of transition between employment and unemployment and between unemployment and employment. Weekly constant probabilities of moving from E to U and U to E of each individual were assumed to be logistic functions dependent on the characteristics of the individual. The parameters of these logistic functions were estimated to maximize the likelihood of the observed spells of employment and unemployment. Because of lack of data, periods out of the labor force were treated, per necessity, as blank. From the parameter estimates and the characteristics of the population, the average estimated weekly probabilities for the whole population of older men of moving from E to U and U to E were .0036 and .1426, respectively.

Simulations of the simple Markov model with all persons having these weekly transition probabilities gives an estimate of the extent to which multiple spells are shorter due to the observation period of one year. Use of the average estimated probabilities of .0036 and .1426, which were for the same time period and the exact same population, make the results dimensionally of the right order of magnitude. There is an added reason for confidence in the order of magnitude from simulations using these estimates:

the average U to E transition probability of .1426 predicts an estimated completed spell length of 7.0; this predicted spell length differs by only 10 percent from the actual completed spell length recorded in Table 4 of 6.3 weeks. Table 5 reports the results of simulations of 100,000 labor force experiences for a population with transition probabilities of .0036 and .1426.

Table 5 shows that, for this population, single spells are approximately 13 percent longer than double spells, and double spells are 23 percent longer than triple spells. This compares with Table 4 in which single spells were 65 percent longer than double spells, and double spells were 50 percent longer than triple-plus spells. Thus the finite observation period bias does not seem capable of explaining the observed difference between single and multiple spells, although it can explain a sizable fraction of the differences in lengths between double and triple spells.

III. Conclusion

This paper has found that the average length of single spells of unemployment is longer than the average length of multiple spells. The conclusion appears robust, being due neither to the underreporting of unemployment experience by teenagers and women, nor to the interruption of uncompleted spells by the beginning or end of the calendar year, nor to the finite period of observation.

Consequently, this paper strongly suggests that statistics on average unemployment durations of completed spells seriously underestimate the unemployment experience of all groups of persons considered in this paper: persons with single spells, because on average single spells are longer than multiple spells; and persons with multiple spells, because the unemployment experience of these persons includes the multiplicity of their spells.

REFERENCES

G. A. Akerlof and B. G. M. Main, "Pitfalls in Markov Modeling of Labor Market Stocks and Flows," *J. Human Resources,* forthcoming 1981.

J. E. Bregger, "Unemployment Statistics and What They Mean," *Mon. Labor Rev.,* Nov. 1971, *16,* 22–29.

K. B. Clark and L. H. Summers, "Labor Force Transitions and Unemployment," mimeo., Cambridge, Mass., Apr. 1978.

M. S. Feldstein, "Lowering the Permanent Rate of Unemployment," U.S. Congress, Joint Economic Committee, Washington 1973.

S. H. Garfinkle, "The Outcome of a Spell of Unemployment," *Mon. Labor Rev.,* Jan. 1977, *22,* 47–54.

R. E. Hall, "Turnover in the Labor Force," *Brookings Papers,* Washington 1972, *3,* 709–56.

H. Kaitz, "Analyzing the Length of Spells of Unemployment," *Mon. Labor Rev.,* Nov. 1970, *93,* 11–20.

R. D. Morgenstern and N. S. Barrett, "The Retrospective Bias in Unemployment Reporting by Sex, Race and Age," *J. Amer. Statist. Assn.,* June 1974, *69,* 355–57.

J. O'Neill, "The Male-Female Unemployment Differential—Findings from Work Experience Survey," U.S. Council of Economic Advisers, Washington, Nov. 21, 1972.

Herbert S. Parnes et al., *The Pre-Retirement Years: A Longitudinal Study of the Labor Market Experience of Men,* Vol. 1, Manpower Research Monograph No. 15, U.S. Department of Labor, Washington 1968.

______ **et al.,** (1972a), *The Pre-Retirement Years: A Longitudinal Study of the Labcr Market Experience of Men,* Vol. 2, Center for Human Resource Research, Ohio State Univ. 1972.

______ **et. al.,** (1972b), *The Pre-Retirement Years: A Longitudinal Study of the Labor Market Experience of Men,* Vol. 3, Center for Human Resource Research, Ohio State Univ. 1972.

G. Perry, "Unemployment Flows in the U.S. Labor Market," *Brookings Papers,* Washington 1972 *2,* 245–78.

S. W. Salant, "Search Theory and Duration Data: A Theory of Sorts," *Quart. J. Econ.,* Feb. 1977, *91,* 39–57.

U.S. Bureau of the Census, *Interviewers' Reference Manual,* Rev., Washington, Feb.

1976.

U.S. Bureau of Labor Statistics, *Special Labor Force Report #127, Work Experience of the Population, 1969,* Washington 1970.

——, *Special Labor Force Report #201, Work Experience of the Population, 1976,* Washington 1977.

——, *Special Labor Force Report #172, Job Tenure of Workers January, 1973,* Washington 1975.

Community Development: The Effects of Growth and Uncertainty

By J. VERNON HENDERSON*

The development of metropolitan area communities that independently finance and provide some or all of their local public services has been carefully explored in the literature. Papers by Charles Tiebout, Bryan Ellickson, Martin McGuire, and Bruce Hamilton establish that, because of uniformity of public good demand and financing considerations, residents of a metropolitan area will tend to stratify into fiscally independent communities where the residents are relatively homogeneous in tastes and wealth. Such stratification may involve the formation of large communities such as suburbs, small communities such as condominiums, or both, depending on the institutional setting and the range of public services under investigation. The stratification result and that type of modeling is the starting point of this paper, from which I will go on to explore two main topics.

First is the dynamic development of these stratified communities, an issue that cannot be explored in the static models in which the stratification result has been derived. Consider either an established community or one just forming where future growth of residents with similar public service demands is expected. Typically, a developer(s) would own the land which future residents would buy, whereas the initial residents would hold zoning and development rights on that undeveloped land plus the power to specify public service levels. Then, in order for land sales and further development to occur in any period, both groups must implicitly or explicitly bargain and agree upon the terms of sales to new residents, given that neither party can proceed unilaterally with development. The agreement will cover lot sizes, housing, and public services to be offered to new entrants and the division of any land sale profits.

Given this bargaining situation, what lot sizes, public services, and prices will be offered to future residents? If there are profits above opportunity cost, what will be the division of profits between existing residents and the developer, and what is the mechanism for the division? Noel Edelson and Michelle White each examined one of these future periods of a community's development in a static framework, but they have done so assuming that either the residents or the developer have complete control and can unilaterally specify the offerings to newcomers and extract all profits. Under that assumption in general as we will see later, the lot sizes and public services offered will not be Pareto efficient. For example, existing residents will act to "overzone" lot sizes to increase the property tax base and the taxes paid by new residents. In this paper, I examine the problem in a setting of bilateral bargaining and exchange between developers and existing residents. The bilateral exchange is examined in a dynamic context where community formation is linked to later growth. Under these conditions it can be shown that in general the bargaining outcome will be to offer later residents Pareto-efficient levels of lots and public services.

The second topic explored in this paper is the type of contractual arrangements that will arise between the developer and initial residents of a community. For example, the process of bilateral exchange may result in future lot sizes and public service levels for

---

*Brown University. William Fischel and Yannis Ioannides provided helpful comments on the earlier version of this paper. I also benefited from discussions with the members of the Urban Economics Workshops at the University of Chicago and Columbia University, the Theory Workshop at the University of Pennsylvania, and a seminar at the University of Washington.

later entrants being contractually fixed at the time of community formation in new communities, or fixed prior to expected growth in established communities. I will examine why and under what conditions this and other types of contractual arrangements arise.

An excellent illustration of the particular arrangement just mentioned is in "new towns" and, to some extent, other large-scale development projects such as some condominiums (see Uriel Reichman). Development in these towns is governed by a combination of land use or zoning regulations, public service infrastructure, and regulations on future public services, all based upon development plans for the town submitted by the developer. Future public service management is governed by a homeowner's association, or a "residential private government," in which the developer can hold a majority of votes up to the first 10–15 years of the community's life. A basic question this paper answers is why people would choose to live in these types of communities and hence agree to contractual arrangements where they forfeit their political rights to enact changes in community development and characteristics for some considerable period of time.

A discussion of contractual arrangements would be incomplete without an analysis of how risks under uncertainty will be shared and shifted between two parties in a bargaining situation (see Joseph Stiglitz, Martin Baily, Steven Shavell, and Costas Azariadis). In the second part of the paper, I will explore how uncertainty about the extent of future demand to live in a community, and hence the extent of possible profits and losses from future land sales and public service provision, will affect the prices of land to initial residents when a community forms. Uncertainty also effects the design and implementation of regulations governing future land uses and public services. As we will see, due to risk aversion, future lot sizes may be set larger or smaller than sizes that would otherwise be characterized as Pareto efficient. Uncertainty also affects the relative efficiency of different types of contractual arrangements.

I. Zoning and Fiscal Policies under Certainty

The topics of this paper are analyzed in a two-period model. In period 1, I examine one group of residents with similar tastes and wealth who wish (can be induced) to form a community for the purpose of jointly consuming local public services. Within any community that forms, it is assumed the land is initially owned by one developer or group of developers who act as one (such as a land development corporation).[1] I will also consider a situation where the community owns the land.

Although within any potential community the land is owned by one agent, residents face a development market for communities that is competitive in the following sense. Our group of people face an array of potential and/or existing communities they can join, some or none of which may offer exactly identical features. Community formation can occur in the context of other communities forming in period 1, with some developers ending up with developments and others not. Regardless, let us assume there are sufficient potential and existing developments so that, in setting prices and designing his potential community, a developer does not perceive that his price-characteristic offerings affect offerings elsewhere—developers do not exhibit strategic pricing behavior.

In period 2, there is population growth in the community because there are additional potential residents who can be induced to join the community at prices and quantities yielding some positive (or, at least nonnegative) benefits to either the developer or the initial community residents, or both. It is presumed that these are people who are similar to initial residents, but who are new to the metropolitan area, or who for some reason had to remain in some other community in period 1. These people face an array of possible existing or potential communities they can join. There may be communities competing for these residents who can make offers that are identical to our community's offer or our community might be able to make a unique offer.

---

[1] The developers will have previously purchased the land from people such as farmers, and the prices they paid may reflect future profits they may earn. I will not investigate this issue, but instead concentrate on what the developers do once they own the land.

This characterization of the same developer being involved in both the formation and growth of a community seems a reasonable characterization of the formation and growth of various new suburban communities in rapidly growing parts of the South, Central, and western United States, and it is certainly a reasonable characterization of the development of new towns and condominiums. In the Northeast, however, growth typically occurs in old established towns located next to a core city. I will discuss later in the paper how our central results may be directly applied to growth of established communities such as these.

A. *The Planned Community Model and Solution*

In this setting of community formation and growth, let us determine current and future public service levels, tax rates, prices, and lot sizes in a world of perfect certainty. In the first situation, assume the developer designs the community by suggesting current lot prices, and current and future lot sizes and public services. In making his suggestions, or choices, he acts to maximize the present value of profits. Potential residents in period 1 shop around in the development market and pick the community that offers them the best package.

I will show that the developer's choices as to lot sizes and public service levels will be Pareto efficient, and that under appropriate contractual arrangements his choices will be the equilibrium solution. Moreover, it will also be shown in Part B of this section that his choices as to lot sizes and public services are *identical* to those that the community, or period 1 residents, would choose if they owned the land themselves and made all decisions unilaterally.

In defining variables I use the following notation to read: $x_j(i)$ refers to the level of variable x for the jth period entrant in period i; $i, j=1,2$. In general, I only use both the i and j designations where necessary. Thus since period 2 entrants consume only in period 2, I use only a subscript for their variables, for example, x_2.

Initial residents maximize intertemporal strictly concave utility functions $V=V(x_1(1), x_1(2), g(1), g(2), l_1)$ where $x_1(i)$ is period i consumption of all other goods, l_1 is lot size, and $g(i)$ is period i consumption of public services. It is assumed that housing services are produced solely with land l_1, and that an initial resident cannot alter his initial lot size in period 2 (and hence can only alter his housing consumption by moving). Secondly, it is assumed that, in any time period, per person consumption of public services is the same for all residents (new or old). (Thus $g(1)$ and $g(2)$ are not subscripted.) Initial residents have a wealth constraint where $W_1=x_1(1)+x_1(2)D+R_1+(t(1)+t(2)D)l_1$. The $x(i)$ have spot prices of 1 and D is a discount factor (the inverse of 1 plus the interest rate), R_1 is the purchase price of a lot in period 1, and $(t(1)+t(2)D)l_1$ are current and future tax liabilities on the lot.

These taxes are determined as follows. Tax and assessment rates per unit of land are the same for all residents and do not in general vary with individual land consumption, although we could specify a variety of other assumptions (see Edelson). Thus where $t(1)$ is the combined assessment price-tax rate prevailing in the community, period 1 tax collections are $N_1t(1)l_1$, where N_1 is the number of period 1 entrants. The public budget must be balanced in each period. Public expenditures in period 1 are $g(1)N_1c$, where c is the unit cost of public services. At this point the nature of these public services must be defined.

In a local context, public services cannot be specified as pure public goods because then it would be inefficient for residents to stratify into different local communities to independently provide these goods. Instead they are a type of congested good, where scale and congestion effects are synthesized on the production side. Then c is the unit cost of providing the same quality of public services to N residents, or of providing gN. As N increases, c may decline. In fact, if g

were a pure public good, $dc/dN=-c/N$, whereas if it is a pure private good at the margin of consumption, $dc/dN=0$.

In this paper it is specifically assumed there is an initial threshold level of provision of public services below which unit costs are prohibitive. This essentially defines a minimum size feasible community and encourages the formation of residential "clubs," or communities to facilitate provision of these goods. I assume N_1 exceeds this threshold size; but unit costs could still decline somewhat as we add population in period 2. In general, I assume scale is sufficient so that g is a private good at the margin of consumption, or $c(1)=c(2)=c$. Given a balanced budget, we know

$$t(1)=g(1)c/l_1 \tag{1}$$

In period 2, taxes are $t(2)(l_1N_1+l_2N_2)$, where N_2 is the number of future entrants and l_2 their lot sizes. Equating taxes with expenditures, $g(2)c(N_1+N_2)$, we have

$$t(2)=g(2)\frac{c(N_1+N_2)}{(l_1N_1+l_2N_2)}\equiv g(2)a/b \tag{2}$$

where $a\equiv c(N_1+N_2)$, and b is the community tax base, $l_1N_1+l_2N_2$.

Initial residents have an alternative equilibrium utility level in their next-best potential community, where $\bar{V}$ is viewed by the developer as being fixed. This reflects the competition assumption made earlier. It always profits the developer to charge prices that reduce the initial residents' utility levels to (or infinitesimally above) their alternative $\bar{V}$. Exactly how $\bar{V}$ is arrived at and what its level implies are extensively analyzed below. For example, it is consistent with a situation where all the developer's future profits are transferred to initial residents via low land prices.

Substituting the consumer's wealth constraint into the utility function for $x_1(2)$, we may write $\bar{V}-V(x_1(1), D^{-1}(W_1-x_1(1)-R_1-(t(1)+t(2)D)l_1), g(1), g(2), l_1)=0$. Given the definition of $t(i)$ in equations (1) and (2), we may then define a function describing willingness to pay for a lot in this community:

$$R_1=R_1\big(x_1(1), g(1), g(2), l_1, l_2; \\ W_1, N_1, N_2, \bar{V}, \ldots\big) \tag{3}$$

$$\partial R_1/\partial m=(\partial V/\partial m)/(-\partial V/\partial R_1)$$

$$m=x_1(1), g(1), g(2), l_1, l_2$$

The existence and properties of this function follow from the implicit function theorem. To illustrate $\partial R_1/\partial m$, note that $\partial R_1/\partial l_2= Dl_1N_2g(2)a/b^2$ and $\partial R_1/\partial l_1=D(\partial V/\partial l_1)/(\partial V/\partial x_1(2))-Dl_2N_2g(2)a/b^2$. This period's lot price R_1 is affected by l_2, since increases in l_2 increase the tax base and payments of future residents for a given $g(2)$. This lowers the period 2 tax rate needed to finance $g(2)$ and hence lowers the future tax liabilities of initial residents. The expression for $\partial R_1/\partial l_2$ indicates the present value of the tax liability reduction and hence how much more initial residents would be willing to pay for a lot in period 1. For $\partial R_1/\partial l_1$, the amount a resident is willing to pay for an additional unit of land is his marginal evaluation of land less his increased tax liabilities in period 2 (note his period 1 tax liabilities for the same $g(1)$ are unchanged since period 1 residents equally divide public service costs and thus $t(1)$ falls by the same percent as l_1 rises).

A potential period 2 entrant has a strictly quasi-concave utility function $v=v(x_2, g(2), l_2)$, where x_2 is consumption of all other goods, l_2 is lot size, and $g(2)$ is per person consumption of public services, which potential entrants view as being fixed (either set in period 1 or determined in period 2 by initial residents). The budget constraint is $y_2-x_2-R_2-t(2)l_2=0$ where y_2 is income and R_2 is the price of a lot in period 2. The equilibrium best alternative utility level in another community is $\bar{v}$, which is viewed as fixed by the developer or community. Hence, as before, from the implicit function $\bar{v}-v(y_2-R_2-t(2)l_2, g(2), l_2)=0$, we may define a willingness-to-pay function that characterizes period 2 entrants:

$$(4) \qquad R_2=R_2(l_2, g(2), l_1; y_2, \bar{v}, N_1, N_2, \ldots)$$

$$\partial R_2/\partial m=(\partial v/\partial m)/(-\partial v/\partial R_2)$$

$$m=l_2, l_1, g(2)$$

where, for example, $\partial R_2/\partial l_1=l_2 N_1 g(2)a/b^2$. An increase in l_1 increases the amount period 2 entrants will pay for a lot because for the same $g(2)$ their tax liabilities fall.

The developer's profits on each sale are revenue, R_1 or R_2, less opportunity costs. Any unused land the developer has can be rented out to, say, farmers at the going agricultural rent p; and it is assumed the developer pays no taxes on land not used for residential purposes. Thus the present value of his opportunity cost on a period 1 sale is $p(1+D)$ and on a period 2 sale is pD, where D is the discount factor.

To close the problem, I must make two more assumptions about the developer's situation. In setting prices, the developer may charge initial and later residents different implicit per period prices (given their different alternatives, $\bar{V}$ and $\bar{v}$). To ensure these solutions are consistent with equilibrium ones, assume that once a developer has sold or let his land, there is no subletting or resale of land in a period so that low-priced residents cannot sell off part of their land to high-priced residents.

Community area can be adjusted at opportunity cost p in period 2 to include any entrants the developer wishes to admit, subject to regulations governing lot sizes. In a certain world, since the developer is omniscient and knows period 2 land sales, community size can be fixed in period 1.

1. *Economic Characteristics of a Solution*

Given the above assumptions, the present value of the developer's profits is simply

$$(5) \qquad P.V.=N_1(R_1-p(1+D)l_1)+DN_2(R_2-pl_2)$$

Maximizing $P.V.$ with respect to lot sizes and public services given equations (3) and (4) and rearranging terms, we get[2]

$$(6) \qquad V_l/V_x=p(1+D)/D$$

$$v_l/v_x=p$$

$$N_1V_g/V_x+N_2v_g/v_x=a\equiv c(N_1+N_2)$$

$$(\partial V/\partial g_1)/(\partial V/\partial x(2))=c/D$$

where $V_l=\partial V/\partial l_1$, $V_x=\partial V/\partial x_1(2)$, $V_g=\partial V/\partial g(2)$, $v_l=\partial v/\partial l_2$, $v_g=\partial v/\partial g_2$, and $v_x=\partial v/\partial x_2$. These conditions state that, given our assumptions, relevant marginal rates of substitution in consumption equal relevant opportunity costs for private goods. In addition, a modified Samuelson condition is satisfied for publicly provided goods. Note this last condition is applied to a publicly provided service which is a congested type of good (and may even be a pure private good at the margin of consumption), where all residents are required to consume the same level of service. Given this specification of the problem, these of course are Pareto-efficiency conditions. (I have maximized the welfare of the developer holding everyone else's ($\bar{v}$ and $\bar{V}$) welfare fixed.)

Why does it pay the developer to set goods at Pareto-efficient levels? As an illustration, examine l_2. *Once into period 2*, the developer would maximize profits $N_2(R_2-pl_2)$ by setting l_2 so that $v_l/v_x=p+\partial(t(2)l_2)/\partial l_2$. Given diminishing marginal rates of substitution in consumption, given $\partial(t(2)l_2)/\partial l_2>0$, and given $g(2)$ fixed, this lot size must be smaller than the one where $v_l/v_x=p$. Once into period 2, the developer would reduce lot sizes to reduce the tax liabilities ($t(2)l_2$) of entrants and hence raise the price these entrants are willing to pay to him. However, from a period 1 perspective, the developer must account for the negative effect of l_2 and $t(2)l_2$ reductions upon R_1, the price initial residents are willing to pay (since these reductions raise their future tax rate). These effects on period 1 and 2 profits of altering l_2 to alter tax liabilities cancel out. Similarly $g(2)$ is set according to a Samuelson-type condition, by trading off gains and losses to period 1 and 2 entrants of $g(2)$ changes, as reflected in the lot prices they are willing to pay.

---

[2] As in my earlier paper, one can arrive at equations (6) by characterizing the equilibrium as one where developers are offering price-quantity tradeoffs (R_1 vs. l_2) to match demand side tradeoffs of initial residents.

2. *Contractual Characteristics of the Planned Community Solution*

Equations (6) will be satisfied only under certain contractual arrangements. Let us specify a set of arrangements that allow equations (6) to be satisfied in equilibrium, and label this solution the *planned community solution*. When initial residents choose to live in the developer's community, they agree to a contract where l_1, l_2 and $g(2)$ are fixed and predetermined to all parties, where they pay a set price R_1, and where the developer is free to charge what he can in period 2, given the fixed l_2 and implicitly fixed period 2 taxes. The level of $g(1)$ can also be fixed in the contract. Alternatively, initial residents can be left to vote for and provide $g(1)$ themselves, since under our assumptions in a competitive (costless) political process they will choose the optimal $g(1)$.

The phrase "agreeing to a contract" can mean one of two things. As in new towns and condominium developments, a contract can literally be signed between developers and each resident that fixes current and future lot sizes and public services. Public services are fixed in practice by period 1 provision of all current and future capacity requirements and regulations setting items such as pupil-teacher ratios, population-park and police ratios, frequencies and standards for garbage disposal and street cleaning, etc. Alternatively, as a new community forms, initial residents can simultaneously vote for current and future public service levels and zoning regulations governing current and future lot sizes, and then grant the developer future development rights under those regulations.

Why must there be these binding contracts for equations (6) to be satisfied? If no contracts and regulations are made and the community forms so that residents hold the development and zoning rights and the developer owns the land, once into period 2 we are in a two-party bargaining situation where the participants are free to try to alter l_2 and $g(2)$ relative to the levels in the equations (6). I call this an *unplanned community situation*.

In this situation, following Edelson, developers will want to sell small lots, reducing l_2 so that $v_l/v_x = p + \partial(t(2)l_2)/\partial l_2$ (see above) which maximizes their remaining profits once into period 2. In contrast, following White, community members will want to zone for large lots to maximize the taxes $(t(2)l_2)$ paid by later entrants and hence reduce their own taxes for any $g(2)$ subject to the constraint that $R_2 - pl_2 \geqslant 0$, so that the developer will still be willing to sell the land. In our modeling of community development neither party can act unilaterally. Moreover, a more efficient solution than those employed by Edelson and White is for the developer and initial residents to cooperate in period 2 in a situation where explicit monetary transfers between them can occur; for example, to set lot sizes where $v_l/v_x = p$ as in equation (6). At $v_l/v_x = p$, in a lot sale, gross revenue over land opportunity cost, or consumer's surplus, is maximized; or, equivalently, the sum of profits plus taxes $(R_2 + t(2)l_2 - pl_2)$ is maximized. Note, however, even if period 2 cooperation yields $v_l/v_x = p$, there is still the bargaining problem of the division of consumer's surplus, or taxes plus profits, between the developer and residents to be solved.

While it is possible for period 2 bargaining in an unplanned community situation to yield Pareto-efficient levels of period 2 variables, it is an inferior solution to the planned community solution, as evaluated in period 1 by developers and initial residents. It is inferior because it creates uncertainty. Initial residents and developers are uncertain as to whether $g(2)$ and l_2 will be set in future bargaining at Pareto-efficient levels and as to future taxes and division of profits. Even if the expected values of $g(2)$, l_2, and profits are those in (6) and (5), by Jensen's inequality, risk-averse residents will be worse off from the uncertainty alone (see Section II where $E[V(\cdot)] < V(l_1^*, g_1^*, E[l_2], E[g_2], E[x_2]) = \bar{V}$). Moreover, risk-avoidance behavior may lead initial residents to move

l_1 away from a Pareto-efficient level (see Part A of Section II). In short, the planned community solution has the advantage that, by shopping in the competitive development market in period 1, initial residents can find a binding contract that eliminates uncertainties and guarantees them their best alternative utility level.

3. *An Illustration of a Solution: A Special Case*

Equation (5) is specified so it may appear that both period 1 and 2 entrants must face restricted alternatives (represented by $\bar{v}$ and $\bar{V}$). However, that appearance would be misleading. It could be that community formation in period 1 is *perfectly* competitive, meaning that initial residents face an array of potential developments that can offer *identical* features, where given competition, some of these developers will end up with no communities at all in period 1. Suppose our community of N_1 residents that does form is the only community forming for this particular wealth-income group in the metropolitan, or submetropolitan, area in period 1. This would occur, for example, if N_1 was large enough to meet threshold community size requirements in public service costs, but $N_1/2$ was too small, so that two communities for N_1-type people could not form.

In period 2, potential residents with similar incomes may face an entirely different choice situation. Suppose there are not enough of them to form a community of threshold size and hence their best alternative is to join a community where the other residents have "substantially" different incomes and lot size and public good demands.[3] This means that if all goods were sold at opportunity cost, period 2 entrants would be willing to pay a premium to live in our community, where the people are similar to them. Thus, given $\bar{v}$, we expect $R_2-pl_2>0$, and the developer will make profits on period 2 land sales. In fact, as a general comment on the solution to (5), in any situation $R_2-pl_2\geqslant 0$ in order for the developer to be induced to carry out his period 2 selling plans.

Given, in period 1, that the various potential developers recognize this future profitable situation, they will compete for the N_1 initial residents so as to snare the development that forms and hence actually get the profits. The key to formation is attracting period 1 entrants. This attraction involves lowering period 1 prices. At the limit such competition would lower the equilibrium R_1 and raise $\bar{V}$ until $P.V.=0$, so that $R_1-p(1+D)l_1=-(R_2-pl_2)N_2D/N_1$ and $R_1-p(1+D)l_1<0$, implying that future land sale profits are transferred to initial residents.[4] Thus specifying $\bar{V}$ does not imply restricted alternatives; and sales to period 1 entrants may be at a profit or a loss. (Finally in terms of the issue of whether the farmers who sold their land to developers originally get some of these profits, they do not if either the developer foresees that $P.V.=0$ or suburban farm land is in sufficient supply that it does not warrant a scarcity premium in its residential use.)

In this illustration, community formation has been perfectly competitive, but established growing communities have had some degree of monopolistic power. This may seem a strong characterization, but it is consistent with the literature. In Hamilton's article communities form in a competitive situation; whereas in White or Edelson they grow in a situation where the array of communities that additional residents face is strongly ordered in terms of public service levels offered (Edelson), so that established communities have some degree of monopolistic power vis-à-vis some potential entrants. Moreover, the monopolistic power characterization has logical appeal. For example, if entrants in both periods 1 and 2 effectively face perfectly competitive situations where second-period entrants can also form their own community at the same opportunity costs, period 1 and 2 entrants would never even be willing to live together unless they had identical public good and housing demands. If they can buy at opportunity costs, neither would experience benefits from joint consumption and there would be no basis for growth within the same community. Thus *the whole intertemporal problem really involves assuming that period 2 entrants face restricted alternatives and hence generally pay premiums above opportunity costs.*

---

[3] Here we also can show that a second community formed out of N_2 plus some N_1 people is inferior to a single community situation due to scale economy problems and the potential capital losses on lot sales of N_1 shifters who must be compensated for these losses if they are to move.

[4] There is a technical problem here. If the pot of future profits is fixed and if the actual number of period 1 entrants (n_1) is set below N_1, the pot is divided among fewer people, benefiting those who actually get in. Then the determination of n_1 becomes a complex bargaining problem, in terms of specifying a stable solution. To avoid this we could assume that $g(1)dc/dN_1<0$ and of sufficient magnitude to make it beneficial to admit all N_1 people to lower fiscal costs despite the increased number of shares in profits. In part this problem arises from truncating a multiperiod model into two periods. We want period 1 to end and period 2 to start at that point in time when the community has sufficient public service scale economies and characteristics to have established its monopolistic position.

However, it should be added that, to ensure equilibrium solutions where later entrants may pay higher effective per period land rents than initial residents, we must assume either that entrants in period 2 are "new to the world" or that, if they are not new, they must accept this type of intertemporal price discrimination. They cannot make contracts in period 1 with the developer or some subset of initial residents and farmers. By not allowing subletting, we generally rule out the most likely arrangement, whereby a future resident would buy a lot in period 1 and sublet it to farmers until period 2. Later in Section II when I introduce uncertainty, ruling out these possible arrangements follows more directly. With uncertainty, in period 1 before the state of nature in period 2 is known, potential future entrants may face a variety of metropolitan and-submetropolitan areas as potential locations for period 2. Then even small transactions costs or costs of making contracts rule out their making arrangements with all the potential communities they might end up in.

4. *Additional Considerations*

There are a number of loose ends that need to be discussed. It has been assumed that period 2 entrants are identical. If they differ and if developers can price discriminate perfectly among types of entrants, then in an equilibrium solution each type of entrant has a land consumption condition of $v_l^i/v_x^i=p$ in (6) and an adjusted Samuelson public service condition applies to the community. If developers cannot price discriminate perfectly for either information or institutional reasons, our solutions are altered in ways discussed in my 1978 paper, pp. 36–38. There remains the question of how many of these differing period 2 entrants should be admitted. The Kuhn-Tucker conditions on this question indicate that the ith type of entrants should be admitted if their gross revenue to the developer and community $(R_i+t(2)l_i)$ exceeds their gross opportunity cost $(pl_i+cg(2))$. Note this condition states that, even if land sales are profitable $(R_i-pl_i>0)$, if fiscal losses $(t(2)l_i-cg(2)<0)$ are sufficiently large, the developer-initial resident contract will not allow sales to the ith type of entrant because the $P.V.$ of profits will fall as R_1 declines (for a given $\bar{V}$) with increases in the ith type of entrant in period 2. This condition also applies for the same reason when residents are identical so that $R_2-pl_2+t(2)l_2-cg(2)\geqslant 0$.

There are two potential problems in achieving the solutions. Both involve how much knowledge it is reasonable to endow on the participants. First in the solutions, the choice variable l_1 is specified on the basis of an opportunity cost of lots whereas the gross price to residents reflects opportunity costs, taxes, and profits. Left on their own to determine lot sizes, period 1 residents would choose to alter lot sizes until V_l/V_x equaled this gross price. To prevent this problem of people attempting to "free ride" by lowering land consumption to reduce tax liabilities, the lot sizes proposed by the developer in period 1 must be fixed at the level satisfying $V_l/V_x=p(1+D)/D$ and residents offered the efficient *all or nothing* choice of gross price and l_1. This is a familiar problem analyzed by Hamilton. In terms of how developers know the lot size where $V_l/V_x=p(1+D)/D$ since it would not be revealed in the housing market by marginal consumer choices, Hamilton's answer would be that, given a costless tâtonnement process, potential developers experiment repeatedly with various $l_1, l_2, g(1), g(2)$ combinations until the optimal combination is revealed and shown experimentally to dominate all other combinations.

The second information problem concerns how the developers in period 1 know the R_2 function of later entrants so as to set l_2 and $g(2)$. This is more difficult to handle conceptually. The R_2 function cannot be revealed experimentally in period 1; and thus a knowledge of R_2 implies a knowledge of future entrants preference functions as well as incomes, alternative living conditions, etc. One possible way around this problem that does not generally affect the results is to assume that developers are risk neutral, perceive some distribution of possible R_2 functions, and act upon expected values (see Section II).

Finally, note that in equation (6) there is no presumption as to whether

$$l_1 \lesseqgtr l_2$$

If it is efficient that $l_1 > l_2$, then period 2 entrants are a *fiscal burden*, because they pay less than average taxes for the community-wide per person provision of $g(2)$. Initial residents would agree to such an arrangement providing $\bar{V}$ was achieved; and the developer would find it profitable to lower l_2 below l_1 to satisfy $v_l/v_x = p$ so as to maximize the $P.V.$ of profits by trading off (as in all situations) $t(2)$ and R_1 reductions against $R_2 - pl_2$ increases.

B. *The Self-Development Model and Solution*

There is another efficient development arrangement, which I label the *self-development solution*, that yields solutions identical to the planned community solution. If initial residents act collectively, they can buy all vacant land from the developer in period 1, set l_2 so $v_l/v_x = p$ in period 2, and split period 2 profits amongst themselves. Then the problem becomes one of initial residents maximizing utility, given a purchase price $\bar{S}$ they *each* pay to the developer for both their own land plus their share in the community ownership of vacant land. The purchase price $\bar{S}$ is defined to be that part of the price above opportunity costs. (By assumption opportunity costs on vacant land can always be recovered.) For their share in vacant land, each initial resident will collect in period 2 a share in land sale profits equal to $N_2(R_2 - pl_2)/N_1$. Thus initial residents seek to maximize

(7)

$$V = V\big(x_1(1), D^{-1}\big(W_1 - x_1(1) - p(1+D)l_1 - cg(1) - Dl_1 g(2)a/b - \bar{S} + DN_2(R_2 - pl_2)/N_1\big), g(1), g(2), l_1\big)$$

R_2 is defined in equation (4) and the first-order conditions for this maximization problem are the same as in equation (6). If $\bar{S}$ is such that $V = \bar{V}$ and equations (6) are satisfied, then all variables have the same values as in the previous maximization problem. The $\bar{S}$ will equal $P.V./N_1$ in equation (5). The presumption is that the developer will charge $\bar{S}$ to initial residents when selling them his land, so as to reduce them to $\bar{V}$. Note in the special case discussed earlier where development is "perfectly" competitive $\bar{S}$ will be zero, or the developer (who could be a group of farmers) holding the land will be forced to sell at opportunity cost.

The only potential efficiency advantage of the planned community solution is that it does not require the community to organize and to actively take over the entrepreneurial role of developer. In the planned community solution, the developer performs all functions and the community has only to monitor compliance with the contract. While the community could simply hire the developer as a manager, they would have to monitor his activities to ensure that not only were quantities set correctly, but that efficient prices were charged. Secondly, considerable group decision making is required.

II. Community Characteristics under Uncertainty

To introduce uncertainty, I assume that the number of period 2 entrants is a random variable $N(\theta)$, whose actual value is determined by the state of nature in the economy and metropolitan area. The community and developer perceive the same distribution function for $N(\theta)$, where for all θ, we assume $0 \leqslant N(\theta) \leqslant N_1$. An uncertain $N(\theta)$ has a direct impact on the period 2 tax rate, or $t(2) = g(2)c(N_1 + N(\theta))/(l_1N_1 + l_2N(\theta))$, so that in period 1 the community's future fiscal position is uncertain. An uncertain $N(\theta)$ also makes future land sale profits $(R_2 - p_2l_2)N(\theta)$ an uncertain magnitude in period 1. Under uncertainty the l_2 and $g(2)$ that are best in period 2 will vary with the state of nature because as $N(\theta)$ varies, the Samuelson condition determining $g(2)$ consumption is altered. As $g(2)$ varies, the optimal l_2 will also vary.

The introduction of uncertainty eliminates any possible equivalence of the planned, unplanned, and self-development situations presented above. The planned situation from Section I involves making choices concerning period 2 lot sizes and public service levels before the state of nature and the optimal l_2, $g(2)$ combination in period 2 are known. Risks are shared, with the community bearing risks on the fiscal side in terms of any fiscal surplus or deficit connected with period 2 entrants $(t(2)l_2 - cg(2))N(\theta)$ and with the developer bearing risks in terms of the magnitude of future profits. In contrast to the planned situation, in the self-development situation in Part B, Section I, period 2 choices do not need to be made until the state of nature and the optimal l_2, $g(2)$ combination are known. However, the community, since it owns all land, bears the risk of both uncertain land sale profits and fiscal surpluses or deficits.

Uncertainty also alters the criteria for choosing lot sizes and public service levels, to include considerations of risk reduction and shifting. In particular, in a planned community situation, the community will tend to either a large or small lot zone to shift risk to the developer.[5]

In the analysis below I will generally assume that the developer is risk neutral and initial community residents are risk averse. ($V(\cdot)$ was assumed in Section I to be strictly concave.) This juxtaposition could reflect either taste differences and the entrepreneurial spirit of the developers, or wealth, shareholder composition, and portfolio differences in the case of large land development companies. It is an assumption typical of the contract literature (see Shavell, Baily, and Azariadis). Let us now turn to a formal analysis and comparison of the different routes for community development.

A. *The Planned Community Solution*

I first examine how the developer's maximization problem is altered under uncertainty. The implicit utility function of residents is altered so that $\bar{V} - E[V(x_1(1), D^{-1}(W_1 - x_1(1) - R_1 - (t(1) + Dt(2))l_1), g(1), g(2), l_1)] = 0$; or substituting in for $t(1)$ and $t(2)$ from equations (1) and (2) $\bar{V} - E[V(x_1(1), D^{-1}(W_1 - x_1(1) - R_1 - cg(1) - Dl_1 g(2)\, a/b), g(1), g(2), l_1)] = 0$, where $a \equiv c(N_1 + N(\theta))$ and $b \equiv l_1N_1 + l_2N(\theta)$. By application of the implicit function theorem there is a function

$$R_1\big(x_1(1), g(1), g(2), l_1, l_2; \bar{V}, W_1, N_1, \vec{N}, \vec{q}, \ldots\big) \tag{8}$$

$$\partial R_1/\partial m = -(\partial E[V(\cdot)]/\partial m)/(\partial E[V(\cdot)]/\partial R_1)$$

$$m = x_1(1), g(1), g(2), l_1, l_2$$

where, for example, $\partial R_1/\partial l_1 = DE[V_l]/E[V_x] - E[V_x Dl_2 N(\theta) g(2) a/b^2]/E[V_x]$; $\vec{N}$ is the vector of $N(\theta)$'s and $\vec{q}$ the corresponding vector of probabilities (alternatively, the N and density functions may be specified as continuous). The R_2 value implied by $\bar{v} - v(y_2 - R_2 - l_2 g(2)a/b, g(2), l_2) = 0$ now varies with the state of nature as $N(\theta)$ varies.[6] Because the variation in $N(\theta)$ varies $t(2)$ and the tax liabilities of period 2 entrants, the amount they will pay to the developer, R_2, for any $\bar{v}$ and l_2 varies. Thus

---

[5] For an analysis of other contracts under uncertainty, see A. Michael Spence and Richard Zeckhauser and Shavell on insurance, Stiglitz on share cropping, and Baily and Azariadis on employment.

(9)

$$R_2 = R_2(g(2), l_1, l_2; y, \bar{v}, N(\theta), \dots) \quad \text{all } \theta$$

$$\partial R_2/\partial m = (\partial v/\partial m)/(-\partial v/\partial R_2),$$

$$m = g(2), l_1, l_2$$

Then, for example, $\partial R_2/\partial l_1 = l_2 N_1 g(2)a/b^2$.

The risk-neutral land developer or land development corporation maximizes the present value of expected profits where

$$(10) \quad P.V. = N_1(R_1 - p(1+D)l_1) + DE[N(\theta)(R_2 - pl_2)]$$

Note this specification assumes either that community land area can be costlessly set sufficiently large in period 1 to accommodate all possible numbers of future entrants, or that period 2 annexation is costless. If community size is fixed in period 1 because there is an increasing opportunity cost of land, the basic impact on the results is to alter any expressions incorporating the opportunity cost of land and to raise the possibility under uncertainty that in some states of nature some future potential entrants will be excluded.

Before proceeding with maximization of (10) under uncertainty, I must first specify the nature of the contract in a planned community situation. I will assume that the developer-resident contract in period 1 fixes l_2 and $g(2)$ at one value regardless of the state of nature in period 2.[7] However, this contract is dominated by two other hypothetically possible contracts. First, l_2 and $g(2)$ could be predetermined but they could be contracted to depend on the state of nature, varying as the actual outcome of $N(\theta)$ varies. Since this is an unconstrained version of our basic solution, it dominates it. I do not specify a state of nature dependent contract as the basic contract, simply because it does not seem to be in practice legally feasible. Zoning laws are not specified so that lot sizes are contingent on the state of nature. However, it is a useful benchmark that will be commented on again in this section and analyzed in footnotes.

The second dominant, hypothetical contract is one where the community avoids risk entirely by contracting out its period 2 tax powers to an insurance company or a (risk-neutral) developer who would then be responsible for providing a predetermined level of public services for a fixed tax rate. Again because of legal and institutional constraints, I do not specify a contract that fixes *both* future taxes and per person public services as the basic contract. That is, in practice the political-fiscal process cannot be entirely removed from the public sector. However, if tax powers could be contracted out to insurance companies or risk-neutral developers, this type of solution dominates any other solution. For example, if l_2 and $g(2)$ are predetermined and fixed as in the basic contract, given the values of l_1, $x_1(1)$, $g(1)$, $g(2)$, and l_2 which are chosen in period 1, then from the wealth constraint in period 2 only $x_1(2)$ varies with $N(\theta)$ as taxes vary with $N(\theta)$. If, in contracting out tax powers, $l_1 t(2)$ is set equal to $l_2 E[t(2)]$, $x_1(2)$ would equal $E[x_1(2)]$; and for the same consumption of other goods, from Jensen's inequality, initial residents would potentially be better off. This would allow the developer to raise R_1 and increase his profits while maintaining $\bar{V}$. If $g(2)$ and l_2 are predetermined but vary with the state of nature, precisely the same arguments apply if $l_1t(2)$ is set equal to $l_1E[t(2)]$, holding l_1, $x_1(1)$, and $g(1)$, and the $g(2)$, l_2, combinations unchanged.[8]

---

[6] It will be generally assumed that period 2 entrants' alternatives represented by $\bar{v}$ do not vary with θ. This assumption is only for expositional purposes and its relaxation is straightforward. If we do allow for this possibility, it is convenient to assume that the variation in $\bar{v}$ with θ is never so great that $R_2 - pl_2 < 0$ for some θ, in which case in period 2 the developers would not admit these residents.

[7] David Cass pointed out to me that an alternative which is possible legally in some communities is to fix $t(2)$ and hence $x_1(2)$, leaving $g(2)$ to vary so that it satisfies the community budget constraint. Depending on the specification of the utility function this could be a better alternative for initial residents. This type of "Proposition 13" possibility is also obviously relevant!

Keeping in mind these two dominant hypothetical contracts, let us proceed with the analysis of a planned community where there is a standard implicit or explicit contract. The size of l_2 is fixed in period 1 at one value by zoning regulations and $g(2)$ also is fixed at one value by regulations or laws setting pupil-teacher ratios, park area-population ratios, capacity requirements, etc. The community's taxes are adjusted in period 2 to balance its budget.

In maximizing (10), because l_2 and $g(2)$ are constrained to one value, for period 2 entrants, x_2 is fixed (given a fixed $\bar{v}$). For period 1 entrants as stated above, only $x_1(2)$ varies because l_1, $x_1(1)$, $g(1)$, and $g(2)$ are all predetermined. First-order conditions yield

$$(11)\quad \frac{E[N(\theta)]l_2}{N_1l_1}(v_l/v_x-p) = -\frac{g(2)l_2}{E[V_x]}E\left[\frac{N(\theta)a}{b^2}(V_x-E[V_x])\right]$$

$$(12)\quad \frac{E[V_l]}{E[V_x]}-\frac{p(1+D)}{D} = \frac{g(2)l_2}{E[V_x]}E\left[\frac{N(\theta)a}{b^2}(V_x-E[V_x])\right]$$

$$(13)\quad v_g/v_xE[N(\theta)]+\frac{E[V_g]}{E[V_x]}N_1 - c\,(N_1+E[N(\theta)]) = \frac{l_1N_1}{E[V_x]}E\left[\frac{a}{b}(V_x-E[V_x])\right]\geqslant 0$$

$$(14)\quad E[\partial V/\partial g(1)]/E[V_x]-c/D=0$$

Starting with equation (11), to evaluate the right-hand side we must evaluate $E[N(\theta)(a/b^2)(V_x-E[V_x])]$. This can be written as $cov[N(\theta)a/b^2, V_x]=E[(N(\theta)a/b^2-E[N(\theta)a/b^2])(V_x-E[V_x])]$. We know that if $N(\theta)a/b^2$ and V_x are both strictly increasing or decreasing functions of $N(\theta)$, the covariance is positive. If one is strictly increasing and the other strictly decreasing, it is negative.[9] $N(\theta)a/b^2$ is a strictly increasing function of $N(\theta)$ (given $N(\theta)\leqslant N_1$ and assuming $l_2<3l_1$). Given concavity, V_x increases (decreases) if period 2 disposable income decreases (increases) with $N(\theta)$, which occurs when $l_1>[<]l_2$. If $l_1>[<]l_2$, period 2 disposable income falls (rises) because future entrants are a fiscal burden (benefit) and increases in their number necessitate increases (decreases) in period 2 tax rates to finance a fixed per person consumption of $g(2)$. Thus if $l_1>(<)l_2$, the covariance is positive (negative); and the right-hand side of (11) is negative (positive).

---

[8] For our argument to hold the insurer must have nonnegative profits. To show that our situation is such that this will be the case, assume the insurer is the developer. The developer's expected period 2 profits are unaffected by offering insurance. Thus we must show that

$$E[N(\theta)(R_2-pl_2)]=E[N(\theta)(R_2^*-pl_2)]$$
$$+E[N_1l_1+l_2N(\theta)]E[t(2)]-E[(N_1l_1+l_2N(\theta))t(2)]$$

The left-hand side is his original expected profits. The right-hand side is his new expected land sale profits plus expected tax revenue minus expected expenditures (given for each θ, $g(2)c(2)(N_1+N(\theta))=(N_1l_1+N(\theta)l_2)(t(2)))$. We know for both equilibria $\bar{v}=v(y-R_2-l_2t(2), g(2), l_2)=v(y-R_2^*-l_2E[t_2], g(2), l_2)$. Therefore, for each $\theta, R_2^*=R_2+l_2t(2)-l_2E[t_2]$. Substituting into the above expression, we see the left-hand side equals the right-hand side.

[9] This may be proved by noting that in the covariance expression since $E[(N(\theta)a/b^2-E[N(\theta)a/b^2])E[V_x]]=0$ we may replace $E[V_x]$ by any constant. Pick the value $V_x(\theta^*)$ which is the value of V_x at the $N(\theta^*)$ where $N(\theta^*)a^*/b^{*2}=E[N(\theta)a/b^2]$. Then both parts of the covariance expression have the same crossover, or zero, $N(\theta)$. Thus if both parts are either increasing or decreasing for each $N(\theta)$, the covariance is the product of two either nonnegative or nonpositive magnitudes, so that the covariance is positive. Otherwise, it is negative.

In summary,

$$(11') \quad v_l/v_x - p \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} 0 \text{ if } l_1 \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} l_2$$

Thus we have the following result, by diminishing marginal rate of substitution properties. Under uncertainty in a planned community situation, for a given $g(2)$, lot sizes of future entrants are not fixed at a level where consumer surpluses extracted are maximized. *If $l_1 < l_2$, lot sizes are reduced, or "underzoned," relative to the level where $v_l/v_x = p$. If $l_1 > l_2$, they are increased or "overzoned."*

The explanation for the policy is as follows. Suppose $l_1 < l_2$, so that future entrants generate a fiscal surplus, and l_2 is reduced relative to the level where $v_l/v_x = p$. Reducing l_2 reduces taxes of entrants and increases R_2 for a given $\bar{v}$. At the same time as l_2 falls, $t(2)$ rises, which decreases R_1 the developer can charge initial residents given $\bar{V}$. Thus in terms of initial residents, by lowering R_1, the developer is increasing current *certain* benefits (represented by $g(2)l_2E[N(\theta)a/b^2]$ on the right-hand side of (11)) while reducing l_2 and future *uncertain* fiscal gains to initial residents (represented by $g(2)l_2E[V_xN(\theta)a/b^2]/E[V_x]$ on the right-hand side of (11)). Since for risk-aversion reasons, initial residents prefer certain to uncertain gains, as l_2 is reduced initially the losses to the developer from reducing R_1 to maintain $E[V(\cdot)] = \bar{V}$ are less than the gains from R_2 increasing. The limiting factor on this initially profitable trade of current for future dollars is that period 2 per person taxes plus profits decline as l_2 moves away from a level where $v_l/v_x = p$.

With respect to l_1, since the right-hand side of (12) is the negative of the right-hand side of (11), we know[10]

$$(12') \quad E[V_l]/E[V_x] - p(1+D)/D \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} 0 \text{ if } l_1 \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} l_2$$

We have the same type of interpretation as to why l_1 is adjusted from a level where the "marginal rate of substitution," measured by $E[V_l]/E[V_x]$, equals $p(1+D)/D$. The lot size l_1 is reduced when $l_1 > l_2$ and increased when $l_1 < l_2$, for $g(2)$ and period 1 consumption levels held fixed. (However, $E[V_l]/E[V_x]$ equals the marginal rate of substitution for a situation where, say, income is random but prices (land taxes) are fixed, so no risk premium is attached to particular goods. Here the marginal rate of substitution is really defined by $E[V_l] - E[V_x(-\partial x_1(2)/\partial l_1)] = 0$.)

To the extent that comparisons are meaningful, we can say that relative to a situation where $E[V_l]/E[V_x] - p(1+D)/D = 0$ and $v_l/v_x = p$, for $l_1 \gtrless l_2$, the effect of introducing tax uncertainty is to reduce the differential between l_1 and l_2, by increasing l_1 and decreasing l_2 when $l_1 < l_2$ and vice versa when $l_1 > l_2$.[11] This reduction in the differential between l_1 and l_2, or "under-" and "over-" zoning of lot sizes, reduces the magnitude of possible future fiscal gains or losses, and hence reduces the risk to initial residents associated with the financing of public services. Also note that this same type of result holds in the hypothetical contract where l_2 and $g(2)$ are predetermined but vary with the state of nature.[12]

---

[10] Suppose that instead of per person consumption of public services (defined to be pure private goods) having to be fixed, total expenditures on a pure public good have to be fixed. Then all additional residents represent a fiscal gain, regardless of lot size, since they contribute something to cover fixed expenditures. In that case it can be proved that $v_l/v_x - p > 0$ and $E[V_l]/E[V_x] - p(1+D)/D < 0$.

[11] For l_1 and l_2 to be set so that $E[V_l]/E[V_x] - p(1+D)/D = 0$ and $V_l/V_x = 0$, we could have $t(2)$ set so that $t(2) = E[g(2)a/b]$ and have had wealth varying "randomly" so that in equilibrium it precisely equals $W_1 + E[l_1g(2)a/b] - l_1g(2)a/b$. Then the budget constraint reads $W_1(\theta) - x_1(1) - x_1(2)D - cg(1) - R_1 - l_1DE[g(2)a/b] = 0$. To examine the effect of introducing risk premiums (i.e., the right-hand side of (12)) into the problem, we differentiate $E[V_l]/E[V_x] - p(1+D)/D = 0$. For small changes in l_1 and l_2 from equation (12) for, say, $l_1 < l_2$ we have, with the introduction of risk premiums, $E[V_x]^{-1}E[V_{ll} - 2p(1+D)/DV_{lx} + (p(1+D)/D)^2V_{xx}]dl_1 < 0$. This uses the fact that from the original budget constraint $\partial x_1(2)/\partial l_1 = -p(1+D)/D$ and $\partial x_1(2)/\partial l_2 = E[N(\theta)(v_l/v_x - p)]/N = 0$. Given concavity of V, $dl_1 > 0$. The analysis of dl_2 and the case of changes in l_1 and l_2 when $l_1 > l_2$ follows the same method.

[12] If the predetermined l_2 and $g(2)$ may vary with $N(\theta)$, the risk premium, and hence the choice of l_2 and

For $g(2)$ consumption, since the tax cost of $g(2)$ is uncertain, $g(2)$ carries a positive risk premium. Thus the developer will set period 2 public services below the level indicated by the relevant Samuelson condition, given the social opportunity cost of public services, lot sizes, and period 1 consumption levels. The fact that the risk premium on $g(2)$ is positive may be demonstrated by noting $E[(a/b)(V_x-E[V_x])]=cov[a/b, V_x]$ and that both terms increase (decrease) with $N(\theta)$ if $l_1>(<)l_2$.[13]

Note that in the discussions of changes in l_1 and l_2 and $g(2)$, consumption of other goods is always held fixed. In my system where the prices of three goods (l_1, l_2, and $g(2)$) change with the introduction of risk premiums, even if utility functions are additive, in general, when considering simultaneous changes in all goods, it is possible to make definitive statements about the absolute direction of change of only one of the three goods. Partly for this reason I do not present other comparative static results connected with changes in the degree of risk aversion or risk. Also the degree of risk aversion is not well defined for a utility function with more than two goods. Finally even if we measure risk as changes in the variance of $N(\theta)$, the impact on the l_1 and l_2 risk premiums is unclear.[14]

I have two final comments. The altering of l_1 and l_2 to shift and reduce risk occurs because public services are financed by property taxes. If user financing was possible, the analysis would not be applicable. Secondly, my results are similar if developers are also risk averse. However, when the developer is risk averse, he prefers certain profits today to uncertain profits tomorrow. Thus l_2 will be overzoned for some values when $l_1<l_2$ as well as when $l_1 \geqslant l_2$ so as to increase future taxes of period 2 entrants and raise the (certain) R_1 initial entrants will pay (see my 1978 paper).

B. *The Self-Development Situation*

In this situation, the community buys all land in period 1 and is thus free to set l_2 and $g(2)$ unilaterally in period 2. Hence the community can wait to set l_2 and $g(2)$ until the state of nature is known. As in Section I, in period 2, prices and quantities are set so as to extract the full consumer's surplus of entrants above land opportunity costs. This surplus is broken into taxes collected and pure profits $N(\theta)(R_2-pl_2)$ which are split equally among initial residents. As explained before, for the right to self-develop the community must pay the developer (or farmers) a sales price $\bar{S}$ to cover expected profits (given $\bar{V}$ and $\bar{v}$).

Then the community's optimization problem becomes to maximize

$$\text{(15)}\quad E\Big[V\Big(x_1(1), D^{-1}\Big(W-x_1(1) -p(1+D)l_1-cg(1)-\bar{S}-Dl_1g(2)a/b +DN(\theta)(R_2-pl_2)/N_1\Big), g(1), g(2), l_1\Big)\Big]$$

where l_2 and $g(2)$ will vary with the state of nature. This yields the following relevant

---

$g(2)$, will vary with the state of nature. For example, for l_2, we have $E[V_x](v_l/v_x-p)=g(2)l_1N_1a/b^2(E[V_x]-V_x)$ for all $N(\theta)$. For each θ, l_2 is altered relative to an l_2 where $v_l/v_x-p=0$ so as to optimally alter R_1 (represented by $g(2)l_1N_1a/b^2E[V_x]$) which is paid before $N(\theta)$ is known, relative to taxes (represented by $g(2)l_1N_1(a/b^2)V_x$) which are assessed after $N(\theta)$ is known.

[13] Not surprisingly, since the risk of holding $g(2)$ is related to the variability of taxes, the risk premium on $g(2)$ is related to the general risk premium π in the problem, where π is defined by $V(E[x_1(2)]-\pi; x_1(1), g(1), g(2), l_1)=E[V(x_1(2); x(1), g(1), g(2), l_1)]$. If π is small and V is approximated by a quadratic, the left-hand side in the definition is commonly approximated by $V-\pi V_x$ and the right-hand side by $V+(V_{xx}/2)\ \mathrm{var}[x_1(2)]$ (see John Pratt). Simplifying and solving we get $\pi=(1/2)(-V_{xx}/V_x)\ \mathrm{var}[l_1g(2)a/b]$. For $g(2)$ the right-hand side of (13) is approximated by $N_1/g(2)(-V_{xx}/V_x)E[l_1^2g(2)^2(a/b-E[a/b])^2]$. Thus, the risk premium on $g(2)$ per initial resident is $2\pi/g(2)$.

[14] On defining risk aversion with many commodities, see Richard Kihlstrom and Leonard Mirman. The problem is that a proper definition of an increase in risk aversion is $V^*=k(V)$ where $V(V^*)$ is the old (new) utility function and $k'>0$, $k''<0$. For more than two goods, this definition does not translate into a simple measure of the degree of risk aversion. With respect to changing the variance of $N(\theta)$, the impact on the l_1 and l_2 risk premiums is unclear since it raises the probability of both low and high fiscal losses (or gains).

first-order conditions

$$\text{(16)}\quad v_l/v_x-p=0 \qquad \text{for all } \theta$$

$$v_g/v_x N(\theta)+V_g/V_x N_1 = c(N_1+N(\theta)) \qquad \text{for all } \theta$$

$$E[V_l]/E[V_x]-p(1+D)/D=0$$

These results have two implications. First as observed above, in contrast to the planned community situation, the maximal amount of consumer's surplus per lot can be extracted in each state of nature as $g(2)$ and l_2 are varied optimally. On the other hand, risk to community members cannot be mitigated and is borne entirely by them, no longer being shared with the developer. Residents now bear the risk of both uncertain future profits and uncertain future tax revenues. Thus while a potential gain in this situation is increased revenues from period 2 entrants, there is a loss from the inability of risk-averse residents to share risk.

In comparing the planned community and self-development situations, one cannot be shown to dominate the other.[15] However, dominance can be demonstrated under particular circumstances.

First, it should be intuitively obvious that if initial residents are risk neutral there are no benefits from risk sharing. Thus there are only losses in the planned community relative to the self-development solution from fixing l_2 and $g(2)$ prior to knowing the state of nature. Therefore, a primary basis under uncertainty for a planned community solution is risk aversion.

Secondly, there is one other particular set of circumstances where self-development definitely dominates the planned community solution. In fact, it can even dominate the planned community solution when l_2 and $g(2)$ are contracted to be dependent on the state of nature. This may be counterintuitive since if l_2 and $g(2)$ can vary with the state of nature in a planned community, planned community solutions should dominate self-development since they involve risk sharing. However, there is one situation where risk sharing is disadvantageous to the community.

Suppose $l_1>l_2$, so that future residents represent a fiscal loss to the community. Of course for these residents to be admitted at all, from Section I, $R_2-pl_2>0$ and $R_2+t(2)l_2-pl_2-cg(2)\geqslant 0$. The community agrees to admit these people because the developer lowers R_1 initially to compensate initial residents. But it means that in period 2 the community faces uncertain losses. However, under self-development since the community collects all period 2 revenue, although entrants represent fiscal losses, their overall profitability $(R_2+t(2)l_2-pl_2-cg(2))$ is positive. Thus self-development involves uncertain gains. In my 1978 paper, it is shown that if the possible fiscal losses are large enough relative to the overall profitability of period 2 entrants, then self-development dominates the planned community solution.

This can be illustrated with a simple example. Suppose there are only two, equally probable, states of nature, where $N(1)=0$ and $N(2)=N_1$. When $N(2)=N_1$, the land sale profit on later entrants above taxes and opportunity costs (a total of $N_1(R_2-pl_2)$) is, say \$$m$ per initial resident; whereas, given $l_1>l_2$ and $N(2)=N_1$, the fiscal loss on later entrants per initial resident $(l_1 g(2)a/b-cg(2))$ is, say, \$$n$ where $m-n>0$. In the planned community case solution given the efficient l_1^*, l_2^*, $g^*(1)$, $g^*(2)$ and $x_1^*(1)$, expected utility is $(1/2)V^*(B;\ldots)+(1/2)V^*(B-n;\ldots)$ where $B=D^{-1}(W_1-x_1(1)-cg(1)-R_1^*-cg(2)D)$. For the self-development case, suppose we restrict self-development to these values of l_1^*, l_2^*, $g^*(1)$, $g^*(2)$ and $x_1^*(1)$ which obviously biases the comparison against self-development. Secondly, we assume developers are paid an $\bar{S}$ so that they are indifferent between solutions. Thus $\bar{S}=m/2$. Then expected utility is $(1/2)V^0(B-m/2;\ldots)+(1/2)V^0(B+m/2-n;\ldots)$. Doing a Taylor's series expansion about B for both functions it can be shown that

$$E[V^*]-E[V^0]=\frac{1}{4}V_{xx}\left(m\ \left(n-\frac{1}{2}m\ \right)\right)$$

$$\left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 0 \text{ as } \frac{m}{2}\left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} n$$

---

[15] See my 1978 paper for a general comparison.

Thus self-development dominates if possible fiscal losses (n) are large relative to overall profitability of future entrants.

C. *The Unplanned Community*

In the first section of the paper, we saw that an unplanned community solution involved a period 2 two-party bargaining situation between the developer and initial residents over zoning laws, public service levels, and division of taxes and land sale profits. It was noted that even if explicit costless transfers are possible between the two parties, which may be quite unlikely, it can be argued that the bargaining solution will set variables at Pareto-efficient levels as in equation (6). But there is still the problem of dividing up the consumers' supluses extracted from future entrants. This problem of uncertain division of uncertain profits derived from variables whose level may be uncertain, even given the state of nature, suggests that an unplanned community would not be a dominant situation for risk-averse people. However, an unplanned community solution does allow l_2 and $g(2)$ to vary with the state of nature (unlike the planned solution) and does involve risk sharing with the developer (unlike self-development). Therefore, it is not really possible to make definitive general statements concerning which is the dominant situation.

III. Extensions

In this paper I have investigated the types and specifications of contracts and arrangements that are likely to arise in new communities. The results are also applicable to established communities. Here we would model how existing residents deal with potential growth in their population, in terms of setting zoning laws and selling development rights. If it is assumed that the development rights can be sold, with the developer making payment in kind in terms of providing public services for new residents (roads, sewers, sidewalks, etc.) and services for old residents (parks, shopping centers, etc.), we have a transfer mechanism between the developer and old residents. A well-defined cooperative bargaining situation could result in lot sizes and public services being set at optimal consumption levels; and the bargain would determine the division of profits that are extracted from new residents, or the price the developer must pay for development rights.

The question is whether the sale of zoning laws and development rights will occur before or after the state of nature is known. As before, this involves two main considerations. The argument for predetermination is that, if residents are risk averse and developers effectively risk neutral, predetermination will enable initial residents to sell development rights for some percent of expected future profits before the state of nature is known and avoid the risk connected with variations in future profits. The argument against predetermination is that it involves fixing lot sizes and future public services before the state of nature and their optimal combinations are known.

REFERENCES

C. Azariadis, "Implicit Contracts and Underemployment Equilibria," *J. Polit. Econ.*, Dec. 1975, *83*, 1183–202.

M. N. Baily, "Wages and Employment Under Uncertain Demand," *Rev. Econ. Stud.*, Jan. 1974, *41*, 37–50.

N. M. Edelson, "The Developer's Problem," *J. Urban. Econ.*, Oct. 1975, *2*, 349–65.

B. Ellickson, "Jurisdictional Fragmentation and Residential Choice," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 334–39.

B. W. Hamilton, "Zoning and Property Taxation in a System of Local Governments," *Urban Stud.*, June 1975, *12*, 205–11.

J. Vernon Henderson, *Economic Theory and the Cities*, New York 1977.

______, "Suburban Community Development: The Effects of Growth and Uncertainty," work. paper no. 77–33, Brown Univ. 1978.

R. E. Kihlstrom and L. J. Mirman, "Risk Aversion With Many Commodities," *J. Econ. Theory*, July 1974, *8*, 361–88.

M. McGuire, "Group Segregation and Optimal Jurisdictions," *J. Polit. Econ.*, Jan./Feb. 1974, *82*, 112–32.

Edwin S. Mills, *Studies in the Structure of the*

Urban Economy, Baltimore 1972.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan. 1964, *32*, 122–36.

U. Reichman, "Residential Private Governments," *Univ. Chicago Law Rev.*, Winter 1976, *43*, 253–306.

S. Shavell, "Sharing Risks of Deferred Payment," *J. Polit. Econ.*, Feb. 1976, *84*, 161–68.

A. M. Spence and R. Zeckhauser, "Insurance, Information, and Individual Action," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 380–87.

J. Stiglitz, "Incentives and Risk Sharing in Sharecropping," *Rev. Econ. Stud.*, Apr. 1974, *41*, 219–56.

C. Tiebout, "A Pure Theory of Local Public Expenditures," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

M. J. White, "Fiscal Zoning in Fragmented Metropolitan Areas," in Edwin S. Mills and Wallace E. Oates, eds., *Fiscal Zoning and Land Use*, Lexington 1975, 175–202.

Racial Trends in Male Labor Force Participation

By DONALD O. PARSONS*

In recent decades, the labor force participation rate of black adult males has fallen sharply. Over the period 1954–76, the rate of nonparticipation of black males 45–54 years old has more than doubled, increasing from 6.8 to 16.6 percent. Among black males 25–44, the nonparticipation rate rose over the same period from 6 to almost 10 percent. The nonparticipation rate for white males also rose systematically, but in much smaller increments. Among whites, nonparticipation rose from 3.2 to 7.5 percent for those 45–54; from 2 to 4 percent for those 25–44. Trends in male nonparticipation in the United States by race and age are reported in detail in Table 1.

Beyond generating general social concern, this large exodus of black males from the labor force has negative implications for the apparent convergence of market earnings by race.[1] The recent findings of Smith and Welch (1978) that the ratio of black to white wage and salary earnings *among those who are in the labor force* has risen from .59 in 1955 to .73 in 1975 is sharply reduced if adjusted for the adverse trend in black labor force participation.[2] This adjustment converts earnings per worker into earnings per capita, potentially a more meaningful measure of the ability of blacks to earn income in the market. Indeed, adjusting for the differential trends in labor force participation, the rise in the black-white earnings ratio is reduced by almost 50 percent.[3] The trend also aggravates earnings inequality *within* racial groups as the fraction of blacks who receive no earnings and the average earnings of those who receive some earnings both continue to rise.[4]

Identification of the source of the growing difference in black and white labor force participation rates is clearly of considerable social importance. Are the large differences in market behavior among black and white males the result of relatively straightforward differences in economic incentive between the two groups, or the result of deeper, or at least less well understood, social forces? My 1980 paper has estimated substantial disincentive effects of Social Security disability benefit levels on male labor force participation behavior.[5] In this paper, I explore the hypothesis that the larger withdrawal of

---

*Professor of economics and research associate, Center for Human Resource Research, Ohio State University. This research was supported by the Employment and Training Administration, U.S. Department of Labor; the National Center for Health Services Research, U.S. Department of Health, Education and Welfare; and the Ohio State University Computer Center. James Brown and Stephen Cook provided excellent research assistance.

[1] For other analyses of trends in the ratio of black to white earnings, see James Gwartney, Richard Freeman, Finis Welch, and James Smith and Welch (1977). For a review of racial differences in wages at a point in time, see the literature cited by Ray Marshall and by Stanley Masters.

[2] The two labor force participation concepts (Smith-Welch and the Bureau of Labor Statistics (*BLS*)) are not identical since Smith-Welch define "workers" as anyone who worked last year while the *BLS* uses a survey week concept. For older workers, at least, the differences are small, since about 80 percent of all workers out during the survey week were out during the entire preceding year. See my 1979b paper.

[3] Specifically 44 percent. The civilian labor force participation rate of males 16 years of age or more dropped from 85.0 to 71.5 percent for blacks, and from 85.4 to 78.7 percent for whites over the period 1955-75.

[4] Smith and Welch (forthcoming) analyze equality of earnings within racial groups for males in the labor force. The inequality discussed here, it should be stressed, is of earnings, not income. Since public transfers have been rising over this period, the effect on income of this participation trend is obviously attenuated. The bias toward earnings increases among highly schooled blacks, noted by Smith and Welch over the period 1960–70, also works toward greater inequality of earnings. The schooling bias in the black-white earnings ratio is less obvious in data Smith and Welch present for the period 1967–74.

[5] See Peter Steiner and Robert Dorfman, Richard Barfield and James Morgan, Michael Boskin, Boskin and Michael Hurd, Richard Burkhauser, and my 1978 paper for analyses of the related retirement decision.

TABLE 1—RATES OF MALE NONPARTICIPATION IN THE CIVILIAN LABOR FORCE BY AGE AND RACE, 1954–76

	25–34		35–44		45–54	
Year	Black	White	Black	White	Black	White
1954	3.7	2.5	3.4	1.8	6.8	3.2
1955	4.2	2.2	3.8	1.7	5.8	3.3
1956	3.8	2.6	3.8	1.9	5.6	3.2
1957	3.9	2.8	3.5	2.0	6.5	3.4
1958	3.7	2.8	3.6	2.0	6.1	3.4
1959	3.7	2.5	4.2	2.0	7.2	3.7
1960	3.8	2.3	4.5	2.1	7.7	3.9
1961	4.1	2.3	5.2	2.1	7.7	4.1
1962	4.7	2.6	5.5	2.1	7.8	4.0
1963	5.1	2.6	5.1	2.2	8.9	3.8
1964	4.1	2.5	5.6	2.4	8.4	3.9
1965	4.3	2.6	5.8	2.3	8.0	4.1
1966	4.5	2.5	5.9	2.4	9.3	4.2
1967	4.5	2.5	6.4	2.3	8.7	4.4
1968	5.0	2.8	6.6	2.4	9.9	4.6
1969	5.6	3.0	7.3	2.6	10.5	4.9
1970	6.3	3.3	6.8	2.7	11.8	5.1
1971	7.1	3.7	8.0	3.0	13.1	5.3
1972	7.3	4.0	8.6	3.0	13.9	6.0
1973	8.3	3.7	8.7	3.2	12.0	6.5
1974	7.7	3.7	9.1	3.3	15.3	7.0
1975	8.6	4.2	10.0	3.6	15.4	7.1
1976	9.4	4.1	9.4	4.0	16.6	7.5

Sources: U.S. Bureau of Labor Statistics, *Handbook of Labor Statistics*, Table 4, 1975, and Table 4, 1977.

blacks results from the differentially large effect of the increasingly generous and progressive structure of Social Security benefits on workers with the least favorable labor market opportunities.

This paper proceeds as follows. In Section I, the interrelationship between labor force participation and the Social Security disability program is discussed. Then in Section II, a model of labor force participation is estimated using data from the National Longitudinal Surveys of older males. This allows an examination of racial differences in the determinants of labor force participation. Finally in Section III, the cross-sectional model is put to the difficult test of "predicting" labor force participation of prime age males by race over the period 1958–76.

I. The Labor Force Participation Decision

The sharp decline in adult male labor force participation among both blacks and whites has coincided with a rapid expansion of the Social Security disability program. From its creation in mid-1957, the program has grown to the point that almost 3 million disabled workers were paid Social Security benefits in an average month in 1977. Applications and enrollments in the program increased rapidly in the last decade. Given the absence of evidence on any corresponding increase in the incidence of poor health in the population, the expansion of the system must result from a social decision. In this section I discuss a number of economic factors which will influence the (simultaneous) decision to leave the labor force and apply for Social Security disability benefits.[6]

The severe restrictions placed on qualification for Social Security disability benefits would seem to eliminate choice. The definition of a qualifying disability is "the inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months" (Social Security Administration, p. 98). The difficulty of accurate diagnosis of health status, however, particularly when the subject is potentially uncooperative or even dissembling, apparently injects considerable uncertainty into the qualification process and introduces the possibility that unqualified individuals will be accepted into the program.

---

[6] A more formal model can be found in my 1980 paper.

From the individual viewpoint, the principal cost of the program is the foregone earnings of any "substantial gainful employment" he might otherwise obtain. In the early 1960's, earning as little as $100 per month was evidence of the capability to undertake substantial gainful employment. Monthly earnings of $50 to $100 were held to be ambiguous in that regard, requiring further study before eligibility could be established. Since there is a five-month waiting period before benefits could begin (six months during the sample period of the data analyzed below), the applicant must incur a significant period without (observable) earnings even should benefit status ultimately not be granted. The probability of application denial is not small. In 1975, 60 percent of all "initial determinations" on disability applications were denials.[7]

The individual's decision to seek Social Security benefits is a gamble in which the applicant foregoes market earnings in order to obtain a chance to acquire rights to a flow of future welfare payments under Social Security. The higher the ratio of Social Security benefits to market wages, the more attractive the gamble and presumably the greater the likelihood the individual will choose to play. The opportunity to earn unreported income also increases the likelihood of applying for disability benefits and correspondingly reduces measured, if not actual, labor force participation.

Other factors also influence the appeal of the Social Security program. The worse the individual's health condition, the more likely is approval of the application and again the more attractive the gamble. The laxity or strictness of the qualification screen will also influence the expected returns to applying for benefits. Howard Marvel reports on evidence of the importance of denial probabilities on application rates.

The Social Security benefit schedule is sharply progressive. The level of basic benefits per month in 1966 (the date of the sample) was $44 per month for all disabled workers whose covered monthly earnings (computed under a complex, historically based formula) were less than $70. Benefits rose by 63¢ per month for each $1.00 in average monthly earnings between $70 and $110. For each dollar of earnings above $110 until the covered earnings limit was reached ($400), benefits increased by 23¢. Market earnings above the Social Security tax ceiling yielded no additional benefits. To the extent labor force participation is affected by the ratio of Social Security benefits to market wage rates, demographic groups such as black males with low average market earnings will be differentially affected by the progressive benefit structure of the Social Security disability program. Whether this critical difference in economic position is capable of explaining all or even a substantial portion of the difference of black and white male labor force participation remains to be established.

II. Race and Labor Force Participation: A Cross-Sectional Model

In this section I report on a cross-sectional analysis of the determinants of labor force participation which provides considerable insight into the sources of racial differences in this critical economic variable. The data set which will be used to estimate the labor force participation model is the older men's sample of the National Longitudinal Surveys (*NLS*), a ten-year panel survey of males 45–59 years of age in 1966, the initial year of the survey.[8] This national survey of

---

[7] See Ralph Treitel.

[8] The study by Herbert Parnes et al. includes an extensive discussion of the data, including survey technique and an overview of the results.

5,000 individuals systematically oversampled black households so that approximately 30 percent of the sample is black. Sample means and standard deviations of important variables are reported in the Appendix.

The discussion of the labor force participation model in the last section suggests that economic factors as well as health conditions will influence the labor force participation decisions of both blacks and whites. The economic factors included in the model are the Social Security replacement ratio (the ratio of potential benefits to market wage rate), the ratio of general welfare availability in the local area (states) to the individual's wage rate, and the fraction of weeks the worker was unemployed in 1965. Since individuals currently out of the market are less likely to report a current or last wage rate or respond to a question on unemployment last year, wage and unemployment functions were estimated and a predicted value assigned if the variable was not reported. The precise structure of these functions can be found in a data appendix available from the author. The wage rate is a function of age, schooling, occupation of last job, race, and region; the unemployment index is principally a function of occupation and industry of last job. Efforts to correct the wage estimates for censoring bias using procedures suggested by James Heckman indicated that censoring was not a problem—perhaps because of the relatively small part of the sample which was censored. The potential Social Security benefit measure is constructed from the wage measure, using standard Social Security benefit tables (see the Appendix below for details).

The measurement of health condition is a difficult but important task. As I have demonstrated in my 1979b paper, self-reported health measures common to most large-scale labor market surveys are seriously biased for this group since an otherwise healthy individual who decides to leave the labor market has an incentive to declare himself in poor health in order to be eligible for consideration for disability benefits. The panel aspect of the *NLS* provides, however, an indirect measure which *is* objective, mortality experience subsequent to the period of analysis. In particular, mortality dummies are introduced into the model of labor force participation for year of death if the respondent died during a particular interval between interview dates.

In column (1) of Table 2, probit estimates of the likelihood that an individual is in the labor force in the survey week in 1966 are reported. The health variables have a powerful effect on labor force participation. Even individuals who die seven to nine years after the sample period (*MORT* 73–75) have a significantly lower labor force participation rate than do those who survive for at least ten years (the base group). Those individuals who die earlier than seven to nine years have sharply lower participation probabilities.

The estimated coefficients of the economic variables all have the expected signs and are statistically significant. The Social Security measure is powerfully and significantly negative. At the point of means of the independent variables, the elasticity of nonparticipation with respect to the Social Security replacement ratio is 1.8. A 10 percent increase in potential benefits relative to the wage rate will induce an 18 percent increase in nonparticipation. Both the level of general welfare payments and the likelihood of being unemployed have statistically significant negative effects on labor force participation.

Perhaps most importantly, the coefficient of the racial dummy (one if black, zero otherwise) is not only statistically insignificant, it is virtually zero in magnitude. The gross difference in participation in this sample between races is 4 percent (94.5 percent for whites, 90.4 percent for blacks), a highly significant difference given the large sample size. Controlling for economic circumstance and health conditions, however, apparently eliminates any systematic difference in behavior between the two racial groups.

For convenience in later estimation and projection, a single mortality index is then constructed using a weighted average of these seven dummies where the weights are the coefficients of the dummies in the model, normalized by the coefficient on the labor force participation effect of death between

Table 2—The Labor Force Participation of Males 45–59 in 1966, Probit Estimates

	(1)	(2)	(3)	(4)
Constant	4.725	4.725	3.918	3.336
	(12.23)	(12.24)	(10.81)	(7.37)
Black[a]	0.008	0.008	−0.255	0.194
	(0.11)	(0.11)	(4.20)	(2.54)
Price Variables				
SSB/W[b]	−0.018	−0.0180		−0.0104
	(7.14)	(7.17)		(3.66)
WEL/W[c]	−0.0049	−0.0049		−0.0045
	(1.99)	(1.99)		(1.74)
UR[d]	−0.911	−0.911		−0.723
	(5.25)	(5.26)		(4.00)
Demographic				
MORT 66–67[e]	−1.124			
	(6.31)			
MORT 67–68	−0.737			
	(4.12)			
MORT 68–69	−0.869			
	(5.80)			
MORT 69–71	−0.765			
	(6.15)			
MORT 71–73	−0.656			
	(5.18)			
MORT 73–75	−0.346			
	(2.51)			
MORT 75–76	−0.126			
	(0.63)			
Mortality[f]		−1.124	−1.153	−1.060
		(11.24)	(11.84)	(10.20)
AGE[g]	−0.0393	−0.0393	−0.041	−0.0345
	(5.52)	(5.53)	(6.03)	(4.65)
SCH[h]				0.047
				(4.36)
MARRIED[i]				0.388
				(2.71)
*MARRIED*SCHW*[j]				0.0118
				(0.88)
BLUDUM[k]				−0.170
				(2.16)
log of Likelihood	−1022	−1022	−1078	−955
Sample Size	4831	4831	4831	4711

Source: National Longitudinal Surveys. *Note*: Absolute values of asymptotic t-ratios in parentheses. Dependent variable is one if respondent was in labor force in survey week, zero otherwise.

[a] Dummy equal to one if race black, zero otherwise.

[b] The ratio of potential Social Security benefits to the hourly wage rate (see the Appendix).

[c] The ratio of local welfare to the hourly wage rate (see the Appendix).

[d] Fraction of year unemployed in 1965 (see the Appendix).

[e] Dummy equal to one if respondent died between 1966 and 1967 surveys, zero otherwise.

[f] An index of poor health constructed from mortality coefficients in col. (1) with coefficients normalized by coefficient of mortality in first interval (*MORT* 66–67). See text and also fn. 9.

[g] Age in years.

[h] Years of schooling completed.

[i] Dummy equal to one if married, spouse present, zero otherwise.

[j] Years of schooling completed by wife times *MARRIED*.

[k] Dummy equal to one if occupation is blue collar.

1966 and 1967, the first (attempted) reinterview date.[9] The model is reestimated using the single mortality index in column (2), Table 2. Because of the nature of the mortality index construction, the coefficients estimated for all variables other than health will be unchanged (except for incidental rounding error in the index). The model reported in column (2) will be referred to below as the price model.

The question arises whether the gross differences in labor force participation are solely the result of racial differences in health status or are, in fact, induced by differences in the economic circumstances of the two groups. In Table 2, column (3), I report probit estimates of labor force participation, excluding the price variables (Social Security benefits, welfare, and unemployment experience). Both health (as measured by the mortality index) and age have powerful negative effects on labor force participation. Controlling for these factors, however, does not eliminate the systematic racial difference in behavior. The black coefficient remains negative and strongly significant. At the mean values of the independent variables, the state of being black still involves a reduction in labor force participation of 2.3 percent.

In Table 2, column (4), I report on probit estimates of the price model, augmented by a number of socioeconomic variables, including marital status, schooling of respondent and of his wife (if married), and a dummy if the individual's current or last job was manual (blue collar). Controlling for these factors results in a *positive* and significant estimate of the black coefficient. Black adult males are more likely than white to be in the labor force, *ceteris paribus*. Participation is higher if the individual is highly schooled and married with spouse present, and lower if he is (or was) in a manual occupation. Since older blacks are overrepresented in the low-schooling, unmarried, and manual occupation groups, the reason for the reversal is transparent. Controlling for these additional demographic variables reduces somewhat the estimated Social Security effect. The estimated elasticity of nonparticipation with respect to the Social Security to wage ratio is 1.0 at the point of means of the independent variables.

TABLE 3—LABOR FORCE PARTICIPATION OF MALES, 45–59 IN 1966, BY RACE PROBIT ESTIMATES

	Black 1	White 2
Constant	4.264	4.925
	(6.47)	(10.16)
Price Variables		
SSB/W	−0.0183	−0.0197
	(4.41)	(6.15)
WEL/W	−0.0168	0.0164
	(4.71)	(2.62)
UR	−0.304	−1.336
	(1.12)	(5.78)
Demographic		
Age	−0.0275	−0.0450
	(2.27)	(5.06)
Mortality	−1.194	−1.093
	(7.33)	(8.53)
Log of Likelihood	−365	−639
Sample Size	1385	3446

Source: See Table 2. *Note*: Absolute values of asymptotic *t*-ratios are in parentheses. Variables are defined in Table 2.

The important finding that average racial differences in labor force behavior are eliminated when differences in the economic circumstances of the two groups are recognized bears further attention. To determine whether the responsiveness of blacks and whites to the economic variables is also the same, the price model (Table 2, col. (2)) was reestimated separately for blacks and whites. The results are reported in Table 3, columns 1 (blacks) and 2 (whites).

Although differences exist, the similarity of the two models is clear. The important Social Security coefficients are quite close,

---

[9] Specifically the mortality index is

$$\begin{aligned} \mathit{Mortality} = {} & \mathit{MORT}\,66\text{-}67 + 0.656\,\mathit{MORT}\,67\text{-}68 \\ & + 0.773\,\mathit{MORT}\,68\text{-}69 + 0.681\,\mathit{MORT}\,69\text{-}71 \\ & + 0.584\,\mathit{MORT}\,71\text{-}73 + 0.308\,\mathit{MORT}\,73\text{-}75 \\ & + 0.112\,\mathit{MORT}\,75\text{-}76 \end{aligned}$$

where *MORT* is a dummy equal to one if the individual died between the two reference dates, zero otherwise.

-0.0183 and -0.0197 for blacks and whites, respectively. Racially separate estimates of the demographic model (Table 2, column (4)), not reported here, yield a similar finding (-0.011 and -0.0107 for blacks and whites, respectively).

The age and health coefficients are also similar. The unemployment effect on labor force participation, however, is substantially lower for blacks. The reason this is so is not clear. The only coefficient to differ qualitatively is the welfare index coefficient. While strongly negative, as expected, for blacks, the coefficient is not only positive but statistically significant for whites. That blacks should be more sensitive to welfare levels is moderately easy to believe (it may simply reflect a curvilinearity in effect); that whites should respond positively is more difficult to understand.

The overall impression, however, is one of similarity and not difference. In the next section I report on projections over time of the labor force participation model shown in Table 2.

III. Trends in Nonparticipation

Since the objective of this study is the explanation of recent increases in nonparticipation in the labor force among prime age males, particularly black males, the question naturally arises whether the model estimated above "predicts" recent trends in nonparticipation. In this section, I will project expected nonparticipation rates for the *NLS* sample over the period 1958–76 using the 1966 estimated labor force participation model and then compare these projections with actual labor force behavior by this group. The correspondence, or lack thereof, of the projections and actual behavior will provide a powerful test of the validity of the underlying model.

The analysis will focus on projecting the trend in nonparticipation of males 45–54 in the sample and comparing that series with the *BLS* series for blacks and whites 45–54. The year 1958 was the first full year of the Social Security disability program. The disability program became available to workers 50–64 in mid-1957; to workers of all ages in late 1960.[10]

The ability to "explain" time-series trends is, of course, a difficult test for any cross-sectional model. The effects of factors which are constant at any point in time but vary over time cannot be estimated in a cross-sectional analysis. In the present case, the relative severity of the medical determination of disability under Social Security is one potentially important factor that may be reasonably uniform at a point in time yet vary significantly over time (see Marvel). Second, the cross-sectional analysis is assumed to be estimated on a sample in long-run equilibrium and therefore provides no information on the adjustment lags of behavior (in this case, labor force participation) behind changes in market conditions.

Time-series projections were constructed over the period 1958–76 in the following fashion. Individuals 45–54 in 1966 were assumed to have behavioral patterns specified by the simple price-health model of labor force participation reported in Table 2, column (2). Among the independent variables, age is held constant since I wish to characterize the behavior of individuals of given age. Trends in the remaining independent variables were constructed in the following way. For mortality, welfare, wage rates, and unemployment, the base year (1966) values were adjusted for each individual by trends in closely related, consistent series. The wage rate for each year for a given individual, for example, is his 1966 wage rate times the ratio of personal per capita income in the year in question to personal per capita income in 1966. Potential Social Security benefits were estimated using the projected wage in the appropriate formula for Social Security benefits during the year. The nonparticipation estimates and the underlying data series are available upon request from the author.

The projected nonparticipation rates for blacks and whites are compared in Figure 1

---

[10] The model fails in the period 1954–57, prior to the Social Security disability program. In particular, white nonparticipation is predicted to be virtually zero.

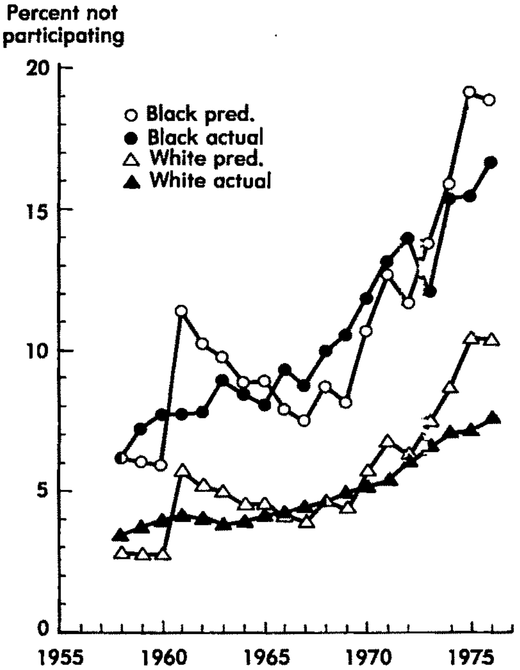


FIGURE 1

with the actual nonparticipation rates. The correspondence is striking. For blacks, the projected rate rises from 6.0 percent in 1958 to 18.8 percent in 1976, the actual rate rises from 6.1 to 16.6 percent. For whites, the projected rate rises from 2.8 to 10.0 percent, the actual from 3.4 to 7.5 percent.

The projections deviate from the actual trend at two different points. The projections jump sharply in 1961, the first year following the expansion of the program to cover workers of all ages (workers 45–49 in this sample). Secondly and importantly, the model overpredicts in the last two years, 1975 and 1976, reflecting the large increases in Social Security benefits relative to market opportunities. This suggests either a lag in response or more likely that the estimates of the Social Security effect are too high. A similar trend analysis using the full demographic model (Table 2, column (4)) underestimates the recent rise in labor force participation. This would indicate that the Social Security coefficients of the two models (with implied nonparticipation elasticities of 1.8 and 1.0) provide reasonable upper and lower bounds for the actual value.

IV. Conclusion

The differentially large decline in labor force participation of prime age males is an obvious source of social concern, both because of its negative effect on the attainment of racial equality of earnings and because of the undesirable consequence on earnings equality within racial groups. Identification of the source of this trend is an obvious first step in determining the broader significance of this trend. In this paper I test the hypothesis that the increasingly generous and increasingly progressive structure of Social Security benefits has affected blacks differentially since they are overrepresented among prime age males with low wage potential.

A multivariate analysis of labor force participation of males 45–59 in 1966 confirms this conjecture. The ratio of potential Social Security benefits to the market wage has a large estimated effect on the participation decision. For the average individual in the sample, a 10 percent increase in benefits without a corresponding rise in wage rates will induce a 10–18 percent rise in nonparticipation. Moreover, controlling for differences in the generosity of welfare relative to market opportunity eliminates any statistically significant differences in the participation rates of black and white males.

To test the validity of the model as an explanation of behavior over time, nonparticipation rates of males are projected over the period 1958–76 using the cross-sectional model. These projections correspond closely with actual rates over this interval. For black males 45–54, actual nonparticipation in this sample rose from 6.1 to 16.6 percent over the period. Using the estimated model in which blacks are distinguished from whites only by their less favorable economic position, the rate would have been expected to rise from 6.0 to 18.8 percent, given trends in Social Security, earning power, etc. For the white sample, nonparticipation rose from 3.4 to 7.5 percent while the estimated model predicted a rise from 2.8 to 10.0 percent.

The conclusion is that the decline in labor force participation, particularly among blacks, is the result of increasingly attractive alternatives to work. The differentially large decline among blacks is due simply to their relatively poor market alternatives and the increasingly progressive structure of Social Security benefits. From a positive viewpoint, one would predict that if wage rates for blacks and whites do ultimately converge, their labor force participation behavior will converge as well.

DATA APPENDIX

A. *Construction of Variables*

Social Security Benefits. This variable represents potential monthly Social Security benefits should the individual qualify as disabled. I use a measure of monthly wages (more precisely hourly wages in 1966 adjusted upward by the growth in personal income per capita times 168 hours) as a measure of the individual's average monthly earnings and use a 1966 benefits table to compute a potential Social Security benefits measure. Since wages have risen over the work lives of these individuals, this wage was adjusted downward over the calculation period for covered wages (1955–66) by the growth in per capita personal income to determine an appropriate covered wage.

Welfare Benefits. The welfare benefits to a nonparticipant in the work force vary widely from locality to locality and state to state. A state index was constructed for 1966 which reflects the generosity of the state's welfare programs. The measure is the number of dollars of public assistance (general assistance plus Aid to Families with Dependent Children) paid in each state per family below the poverty line of $3,000.

Wage Rates and Unemployment. When wage rate and unemployment data were not available for individuals, an estimated value was constructed from regressions which included age, schooling, race, and occupation and industry of last job. The estimated models are available from the author.

B. *Means and Standard Deviations of Variables in the Labor Force Participation Model* (Table 2, cols. (1) and (4)).

Variable	Mean	Standard Deviation
*LFP*1966	0.934	0.248
Black	0.282	0.450
SSB/W	46.50	14.00
WEL/W	11.32	10.51
UR	0.036	0.129
MORT 66-67	0.012	0.110
MORT 67-68	0.014	0.119
MORT 68-69	0.020	0.141
MORT 69-71	0.033	0.178
MORT 71-73	0.032	0.177
MORT 73-75	0.035	0.185
MORT 75-76	0.019	0.138
Mortality	0.0908	0.227
AGE	51.50	4.26
SCH	9.35	3.90
MARRIED	0.874	0.332
*MARRIED*SCHW*	9.02	4.46
BLUDUM	0.633	0.482

REFERENCES

Richard Barfield and James Morgan, *Early Retirement: The Decision and the Experience*, Ann Arbor 1969.

M. Boskin, "Social Security and the Retirement Decision," *Econ. Inquiry*, Jan. 1977, *15*, 1–25.

______ **and M. Hurd,** "The Effects of Social Security on Early Retirement," Nat. Bur. Econ. Res. work. paper no. 204, Stanford 1977.

R. Burkhauser, "The Pension Acceptance Decision of Older Workers," *J. Hum. Resources*, forthcoming.

R. Freeman, "Changes in the Labor Market of Black Americans, 1948–1972," *Brookings Papers*, Washington 1973, *1*, 67–120.

J. Gwartney, "Changes in the Nonwhite/White Income Ratio—1939–67," *Amer. Econ. Rev.*, Dec. 1970, *60*, 872–83.

J. J. Heckman, "Sample Selection Bias as a Specification Error (With an Application to the Estimation of Labor Supply Functions)," rept. no. 7720, Univ. Chicago Center Math. Stud. Bus. Econ., 1977.

R. Marshall, "The Economics of Racial Discrimination: A Survey," *J. Econ. Lit.*, Sept. 1974, *12*, 849–71.

H. P. Marvel, "Disability Applications, Disability Determinations, and Intergovernmental Relations," mimeo., Ohio State University 1980.

Stanley H. Masters, *Black-White Income Differentials*, New York, 1975.

Herbert S. Parnes et al., *The Pre-Retirement Years: A Longitudinal Study of the Labor Market Experience of Men*, Vol. 1 Washington, 1970.

D. O. Parsons, "The Supply and Demand of Public Charity to the Aged," mimeo. 1978.

______, "The Decline in Male Labor Force Participation," *J. Polit. Econ.*, Feb. 1980, *88*, 117–34.

______, (1979a) "Black/White Differences in the Labor Force Participation of Older Males," mimeo., Center for Human Resource Research, Ohio State Univ. 1979.

______, (1979b) "The Male Labor Force Participation Decision: Health, Declared Health, and Economic Incentives," mimeo. 1979.

J. P. Smith and F. R. Welch, "Black-White Male Wage Ratios: 1960–70," *Amer. Econ. Rev.*, June 1977, *67*, 323–38.

______ **and** ______, "Race Differences in Earnings: A Survey and New Evidence," R-2295-NSF, Rand Corp., Mar. 1978.

______ **and** ______, "Inequality: Race Differences in the Distribution of Earnings," *Int. Econ. Rev.*, forthcoming.

Peter Steiner and Robert Dorfman, *The Economic Status of the Aged*, Berkeley 1959.

R. Treitel, "Appeal by Denied Disability Claimants," publi. no. (SSA), 76-11858, U.S. Dept. Health, Education, and Welfare 1976.

F. R. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *63*, 893–907.

Social Security Administration, *Social Security Handbook*, 4th ed., Washington 1969.

Center for Human Resources Research, "National Longitudinal Surveys, 1966–69," Ohio State University.

U.S. Bureau of Labor Statistics, *Handbook of Labor Statistics*, Washington 1975; 1977.

Increasing Downside Risk

By C. MENEZES, C. GEISS, AND J. TRESSLER*

It has now been firmly established in the literature that individuals generally avoid situations which offer the potential for substantial gains but which also leave them even slightly vulnerable to losses below some critical level. This kind of behavior, aversion to downside risk, has been ascribed to a disaster avoidance motive. Robert Lanzillotti's study of pricing objectives in large companies, Richard Rosett's work on call options, experimental investigations by C. H. Coombs and D. G. Pruitt, Robert Masson's data on peasant farming in Mexico, and James Mao's (1970b) survey of the attitudes of business executives to investment options, all attest to the significance attached to downside risk in choice situations.[1] Despite this, no general characterization of the notion of downside risk has been provided in the literature. Prevailing studies have focused mainly on two kinds of measures of downside risk. Probability loss measures, sometimes referred to as lower partial moments, formulate downside risk as a probability-weighted function of deviations below some critical outcome, while skewness measures use the third moment of a distribution about the mean or another predetermined target.[2] The basic limitation of these formulations is that they lack a satisfactory choice-theoretic foundation and fail to distinguish increasing downside risk from other characteristics of distributions, for example, riskiness.

These studies and measures suggest that the intuitive notion of downside risk is concerned with the "placement" of risk in a distribution. One distribution is said to have more downside risk than another if it has more dispersion below a specific target or if it is more skewed to the left. The general notion of a pure increase in risk involves the spreading of probability weight from the center to the tails of a distribution, and conversely, a decrease in risk would result from a contraction of probability weight. In this paper we provide a general characterization of increasing downside risk by combining a spread and contraction of probability weight in such a way that together they result in an unambiguous transfer of risk from the right to the left of a distribution. Such a transfer does not change the overall "riskiness" of the distribution, but simply alters the placement of the risk.

It is natural to expect an individual to be averse to downside risk if he is decreasingly risk averse. Such an individual can be viewed as one whose von Neumann-Morgenstern utility function $u(\cdot)$ has a positive third derivative. We show that one distribution is preferred to another by all downside risk-averse individuals if and only if the distributions differ by probability transfer functions which unambiguously shift risk from the right to the left.[3] This equivalence provides a choice-theoretic foundation for our analysis.

---

*Professor and associate professor, University of Missouri-Columbia, and lecturer, University of Waikato, New Zealand, respectively. We are indebted to M. Proctor, L. Littlejohn, and a referee for valuable comments.

[1] Other examples include Donald Simonson's use of downside risk to explain the behavior of mutual funds and Roger Waud's analysis of the importance of asymmetry in decision makers' preferences for deviations above and below a chosen policy target. For a comprehensive bibliography of the business and psychology literature relating to downside risk, see Robert Libby and Peter Fishburn.

[2] Fishburn's recent paper contains a survey and a general formulation of probability loss measures. Clayton Alderfer and Harold Bierman present experimental evidence of the importance of skewness. Skewness has also been used in discussions on the demand for money (see, for example, John Hicks and S. C. Tsiang), portfolio selection (Giora Hanoch and Haim Levy), and capital asset pricing (Fred Arditti). For a recent treatment of skewness preference, see G. A. Whitmore (1975). Other examples of notions related to downside risk are the early works of A. D. Roy and William Baumol and chance-constraint programming (see Abraham Charnes and William W. Cooper).

[3] An important implication of this result is that it provides a general and intuitive interpretation of the sign of the third derivative of the utility function. The

While these formulations are intuitive and have empirical relevance, they do not provide a practical criterion to rank distributions in terms of increasing downside risk. Accordingly, we formulate a third equivalent criterion which only uses simple properties of distribution functions. This enables us to clarify the relationship between downside risk and stochastic dominance rankings, and to show how existing parametric measures are related to our analysis.

I. Mean-Variance-Preserving Transformations: A Numerical Example

We begin by illustrating how the intuitive notion of increasing downside risk can be expressed in terms of probability transfer functions. In an experimental investigation, Mao (1970b) questioned executives in eight medium and large companies in several industries about their preferences between investment alternatives such as those represented by the following pair of risks:

$f(x)$	$g(x)$
$Pr\{X=1\}=3/4$	$Pr\{X=0\}=1/4$
$Pr\{X=3\}=1/4$	$Pr\{X=2\}=3/4$

He found an unambiguous preference for risks like $f(x)$ over $g(x)$, a finding consistent with the results of many other empirical studies.[4] Since these risks have equal means and variances, preference for one over the other cannot be explained by mean-variance analysis. Mao and others interpreted this preference for $f(x)$ over $g(x)$ as evidence of the importance of downside risk and used skewness or target semivariance to measure it.

The comparative structure of such risk pairs provides the basis for a general characterization of increasing downside risk. For *each* specific value of x, the smaller dispersion of $f(x)$ below this value is exactly matched by its higher dispersion above it; in this sense $f(x)$ can be said to provide better downside protection or greater potential for upside gain. The comparative structure of these risks is explicitly described by the function $t(x)=g(x)-f(x)$. This function is an example of what we call a mean-variance-preserving transformation (*MVPT*). The special properties of $t(x)$ are now illustrated by showing how it can be expressed as a combination of two simpler probability transfer functions, a mean-preserving spread (*MPS*) and a mean-preserving contraction (*MPC*).

Consider the function $s(x)$ shown in Table 1. It is an example of an *MPS*. Adding $s(x)$ to $f(x)$ spreads probability weight on the left of the distribution without changing its expected value. The new probability distribution $h(x)=f(x)+s(x)$ has the same mean as $f(x)$ but is more risky. Now consider the function $c(x)$. It is an example of an *MPC*. Adding $c(x)$ to $h(x)$ contracts probability weight on the right of the distribution without changing its expected value. The probability function $g(x)=h(x)+c(x)$ has the same mean as $h(x)$ but is less risky.

It is apparent that $g(x)$ can be obtained directly from $f(x)$ by a combination of the spread and the contraction since $g(x)=h(x)+c(x)=f(x)+[s(x)+c(x)]$. The means of $f(x)$ and $g(x)$ are the same, and since the variance-increasing effect of the spread is exactly offset by the variance-decreasing effect of the contraction, the variances are also equal. Because the impact of the spread is on the left of the distribution while the contraction affects the right, the net effect of the spread-contraction combination is to transfer dispersion from the right to the left for each value of x. Thus $g(x)$ is skewed to the left while $f(x)$ is skewed to the right. Earlier we used $t(x)=g(x)-f(x)$ to describe the comparative structure of this risk

---

usual justification for a positive third derivative is that it is a necessary (but not sufficient) condition for the risk premium to decrease with increasing wealth (decreasing absolute risk aversion). A *necessary and sufficient* condition for a positive third derivative can be stated in terms of the behavior of the preference intensity function as wealth increases (see D. L. Hanson and Menezes). Here we show the equivalence between the sign of $u'''(x)$ and an individual's preference for or aversion to downside risk.

[4] See Libby and Fishburn, pp. 279–86.

TABLE 1

$X=x$:	0	1	2	3
$f(x)$	0	0.75	0	0.25
$s(x)$	0.25	−0.50	0.25	0
$h(x)$	0.25	0.25	0.25	0.25
$c(x)$	0	−0.25	0.50	−0.25
$g(x)$	0.25	0	0.75	0

pair and called it an *MVPT*. We have shown that $t(x)=s(x)+c(x)$ is also a particular spread-contraction combination which when added to $f(x)$ results in a new distribution $g(x)$ with more downside risk.

Thus the comparative structure of $f(x)$ and $g(x)$ captures the notion of increasing downside risk. To derive a criterion for determining whether one distribution has more downside risk than another, we now analytically formulate the concepts of spreads, contractions and mean-variance-preserving transformations illustrated in this example. This criterion yields a partial ordering on the set of distributions in terms of greater downside risk.

II. Spread-Contraction Combinations and Increasing Downside Risk

In their analysis of increasing risk, Michael Rothschild and Joseph Stiglitz defined a mean-preserving spread as any function which, when added to a probability or density function, transfers weight from the center of the distribution to its tails without changing the mean. The analysis that follows uses a generalization of the specific functional form used by Rothschild and Stiglitz to characterize an *MPS*. Let $f(x)$ be any probability or density function.[5] A function $s(x)$ is an *MPS* if it satisfies the following conditions:

$$(1)\quad f(x)+s(x)\geqslant \text{ almost everywhere}$$

$$(2)\quad \int_0^1 s(x)\,dx=0$$

$$(3)\quad \int_0^1 xs(x)\,dx=0$$

$$(4)\quad s(x)\begin{cases}\leqslant 0 \text{ for } x\in[a,b] \text{ and} \\ \quad <0 \text{ for some subinterval} \\ \quad \text{of } [a,b], \quad a,b\in(0,1) \\ \geqslant 0 \text{ for } x\notin[a,b]\end{cases}$$

The first two conditions simply insure that $s(x)$ is a probability transfer function, i.e., $f(x)+s(x)$ is still a probability function. The third condition indicates that $s(x)$ does not change the mean of $f(x)$. Condition (4) characterizes the essential property of an *MPS*, the transfer of weight from the center to the tails of a distribution.

The converse of an *MPS* is a mean-preserving contraction. We define it as any function which when added to a probability or density function shifts weight from the tails towards the center of a distribution without altering the mean. Formally, if $f(x)$ is a probability function then $c(x)$ is an *MPC* if (1), (2), and (3) are satisfied with $c(x)$ substituted for $s(x)$ and if

$$(5)\quad c(x)\begin{cases}\geqslant 0 \text{ for } x\in[d,e] \text{ and} \\ \quad >0 \text{ for some subinterval} \\ \quad \text{of } [d,e],\ d,e\in(0,1), \\ \leqslant 0 \text{ for } x\notin[d,e]\end{cases}$$

Condition (5) characterizes the essential property of an *MPC*, the transfer of weight from the tails towards the center of a distribution.

The concept of a mean-variance-preserving transformation follows naturally from an appropriate pairing of an *MPS* and an *MPC*. Some additional notation and properties of spreads and contractions are required to define an *MVPT*. Let $S(x)=\int_0^x s(y)\,dy$ and $S^*(x)=\int_0^x S(y)\,dy$. The following properties of $S(x)$ and $S^*(x)$ are easily verified.

$$(6)\quad S(0)=S(1)=0$$

---

[5] Any economically meaningful distribution can be normalized to the unit interval. In what follows we consider distributions defined on [0, 1].

(7) There exists an x_s in (a,b)

such that $S(x)\geqslant 0$ for all $x\leqslant x_s$

and $S(x)\leqslant 0$ for all $x\geqslant x_s$

$$S^*(x)\geqslant 0 \tag{8}$$

$$S^*(0)=S^*(1)=0 \tag{9}$$

Similarly, let $C(x)=\int_0^x c(y)\,dy$ and $C^*(x)=\int_0^x C(y)\,dy$. The following properties of $C(x)$ and $C^*(x)$ are also easy to verify.

$$C(0)=C(1)=0 \tag{6'}$$

(7′) There exists an x_c and (d,e)

such that $C(x)\leqslant 0$ for all $x\leqslant x_c$

and $C(x)\geqslant 0$ for all $x\geqslant x_c$

$$C^*(x)\leqslant 0 \tag{8'}$$

$$C^*(0)=C^*(1)=0 \tag{9'}$$

We are now in a position to formally define an *MVPT*. In the example used in Section I, we paired an *MPS* and an *MPC* in such a way that the effect of the spread occurred before the contraction resulting in an *MVPT* which has the property that when added to a distribution it unambiguously transfers dispersion from the right to the left. This intuitive notion is analytically characterized by the following definition of an *MVPT*. A function $t(x)$ is an *MVPT* if

$$t(x)=s(x)+c(x) \tag{10}$$

$$\int_0^1 T^*(x)\,dx=\int_0^1[S^*(x)+C^*(x)]\,dx=0 \tag{11}$$

$$\int_0^x T^*(y)\,dy\geqslant 0 \text{ for all } x \text{ in } [0,1] \tag{12}$$

and $\quad >0$ for some x in $(0,1)$

The first condition states that $t(x)$ does not alter the mean of a distribution; if $g(x)=f(x)+t(x)$ then $E_g=E_f=\mu$. Given (10), the second condition guarantees that adding $t(x)$ to $f(x)$ does not alter the variance. To show this integrate (11) by parts twice and use properties (6), (6′), (9), and (9′) to obtain

$$\int_0^1 T^*(x)\,dx=1/2\int_0^1 x^2[s(x)+c(x)]\,dx$$
$$=1/2\int_0^1 x^2 t(x)\,dx$$

Since $t(x)$ preserves the mean and $\int_0^1 x^2 t(x)\,dx=0$ by (11), for any constant μ, $\int_0^1(x-\mu)^2 t(x)\,dx=0$. Then as $g(x)=f(x)+t(x)$ and $\mu=E_g=E_f$, it follows that

$$V_g=V_f+\int_0^1(x-\mu)^2 t(x)\,dx=V_f \tag{13}$$

Thus conditions (10) and (11) guarantee that adding $t(x)$ to a probability function does not change the mean and variance.

We now show that condition (12) characterizes the essential properties of an *MVPT*. When written as $\int_0^x S^*(y)\,dy\geqslant -\int_0^x C^*(y)\,dy$ it states that for *each* x the impact of the spread is at least as great as that of the contraction. In this sense risk is transferred from higher to lower values of x, thus increasing downside risk.[6] In conjunction with (10) and (11) it is equivalent to the requirement that an *MVPT* transfer probability weight in the form of dispersion from the right to the left. To derive this interpretation of (12), rewrite condition (11) as $\int_0^x T^*(y)\,dy=-\int_x^1 T^*(y)\,dy$, or equivalently,

$$\int_0^x(x-y)^2 t(y)\,dy=-\int_x^1(x-y)^2 t(y)\,dy \tag{14}$$

Condition (12) guarantees that the left-hand side of (14), and hence its right-hand side, is nonnegative for any x in $[0,1]$. As $g(x)=f(x)+t(x)$ it follows that for each x

$$\int_0^x(x-y)^2[g(y)-f(y)]\,dy$$
$$=\int_x^1(x-y)^2[f(y)-g(y)]\,dy\geqslant 0$$

---

[6] An *MVPT* can also be defined in terms of spread-contraction combination in which the *MPS* comes everywhere before the *MPC*. This makes the leftward transfer of risk even more transparent.

Thus adding $t(x)$ to a probability function $f(x)$ increases the dispersion below x and decreases the dispersion above x by exactly the same amount. This holds for all values of x.

The roles of all three conditions used to define an *MVPT* can be summarized by noting that for each x

$$(15)\quad V_f+(E_f-x)^2=\int_0^x (x-y)^2 f(y)\,dy + \int_x^1 (x-y)^2 f(y)\,dy$$

$$(16)\quad V_g+(E_g-x)^2=\int_0^x (x-y)^2[f(y)+t(y)]\,dy + \int_x^1 (x-y)^2[f(y)+t(y)]\,dy$$

The left-hand sides of (15) and (16) are equal because of conditions (10) and (11); $E_f=E_g$ and V_f and V_g. Subtracting (15) from (16) gives (14), which by (12) is nonnegative for each value of x. We have shown that the conditions which define an *MVPT* create a new distribution which has at least as much dispersion to the left of any value of x, and greater dispersion for some values of x.

These properties of *MVPT*s lead to a general definition of increasing downside risk. It is natural to say that one distribution has more downside risk than another if it can be obtained from the other by a sequence of probability transfers which unambiguously shift dispersion from the right to the left without changing the mean and variance. Such a definition formally captures the intuitive and empirically relevant notions of downside risk suggested in the literature.

DEFINITION:[7] $g(x)$ has more downside risk than $f(x)$ if $g(x)$ can be obtained from $f(x)$ by a sequence of *MVPT*s, i.e., $g(x)=f(x)+\Sigma_i t_i(x)$.

Since it is often difficult to determine whether one distribution can be obtained from another by a sequence of *MVPT*s, it is valuable to have an equivalent criterion to rank prospects using only simple properties of their distribution functions. In the theorem that follows, F and G are the distribution functions corresponding to the prospects f and g.

THEOREM 1: *$g(x)$ has more downside risk than $f(x)$, i.e., $g(x)=f(x)+\Sigma_i t_i(x)$, if and only if*

(i) $E_g=E_f$

(ii) $\int_0^1\int_0^z [G(y)-F(y)]\,dy\,dz=0$

(iii) $\int_0^x\int_0^z [G(y)-F(y)]\,dy\,dz \geqslant 0$ *for all* x *in* $[0,1]$ *and* >0 *for some* x *in* $(0,1)$.

PROOF:

It immediately follows from the properties of *MVPT*s (10), (11), and (12) that if $g(x)=f(x)+\Sigma_i t_i(x)$ then (i), (ii), and (iii) of the theorem are satisfied. The proof of the converse, that (i), (ii), and (iii) imply that $f(x)$ can be obtained from $g(x)$ by a sequence of *MVPT*s, is tedious and requires considerable additional notation. Thus it is omitted.[8]

We conclude this section by noting that the notions of increasing risk and increasing downside risk are conceptually distinct.[9] Rothschild and Stiglitz's general definition of increasing risk implies that if $g(x)$ is riskier than $f(x)$,

$$\int_0^1\int_0^z [G(y)-F(y)]\,dy\,dz>0$$

Thus, in view of (ii) of Theorem 1, $g(x)$ and

---

[7] In what follows the index i is an element of a countable set.

[8] The proof is a constructive process which isolates particular *MVPT*s and subtracts them from $G(x)-F(x)$ until the difference is exhausted. It is available from the authors on request.

[9] This distinction is in contrast to the approach of many authors who try to develop a risk construct which includes downside risk considerations. For example, see Harry Markowitz and Mao (1970a, b). However, the distinction between risk and downside risk is implicit in F. H. Hahn's awkward statement "The 'commonsense' of these results is analogous to that of the familiar distinction between 'gamblers' and 'insurers' even though the utility function is strictly concave" (p. 23).

$f(x)$ are noncomparable in terms of downside risk. Conversely, if $g(x)$ and $f(x)$ are comparable in terms of increasing downside risk, they must be noncomparable in terms of riskiness.

III. Choice-Theoretic Foundations

The results of the previous section show that there exists a transform, an *MVPT*, which unambiguously transfers risk from the right to the left of a distribution. Since the intuitive notion of downside risk is concerned with the placement of risk in a distribution, we argued that it is natural to define an increase in downside risk in terms of a sequence of *MVPT*s. In this section, we provide a choice-theoretic foundation for this formulation of downside risk by characterizing the group of individuals that we would expect to be downside risk averse. We then show that unanimous preference by this group of individuals yields an equivalent definition of increasing downside risk.

As an increase in downside risk involves the transfer of risk leftward in a distribution, we would expect an individual to be made worse off by such a change if he is either decreasingly risk averse or has an increasing preference for risk. For these individuals, increasing downside risk always transfers risk to a range of outcomes over which they are more risk averse, or prefer risk less. The preference-intensity function measures the individual's preference for a surety equal to the expected value of a risk over the risk itself.[10] Stone calls this function the "generalized risk measure" while Milton Friedman and Leonard Savage interpret it as a measure of the disutility (utility for a risk preferrer) attached to a risky prospect. In terms of this function it would be natural to define a downside risk averter as an individual whose preference-intensity function is decreasing. An increase in downside risk shifts risk to those outcomes for which the individual's disutility from risk is higher.[11] Because the preference-intensity function is decreasing if and only if marginal utility is convex, the set of downside risk averters can be characterized as all individuals whose von Neumann-Morgenstern utility functions have a positive third derivative.[12] Let

$$U_3^* = \{u(x) : u'''(x) > 0\} \tag{17}$$

An individual whose utility function belongs to U_3^* is downside risk averse.

We now show that unanimous choice by all downside risk averters induces the same ordering of distributions as one based on a definition of downside risk in terms of spread-contraction combinations. Thus a definition of downside risk based on unanimous choice by all downside risk averters is equivalent to one formulated using *MVPT*s.

THEOREM 2: *$E_f u(x) \geqslant E_g u(x)$ for all $u(x)$ in U_3^* and $F(x) \neq G(x)$ if and only if $g(x)$ has more downside risk than $f(x)$, i.e., $g(x)$ and $f(x)$ satisfy conditions* (i), (ii), *and* (iii) *of Theorem* 1.

PROOF:

Suppose $E_f u(x) \geqslant E_g u(x)$ for all $u \in U_3^*$. We first show that this implies (i), (ii), and (iii). Consider the following pair of utility functions in U_3^*: $u_1 = 1/3\theta x^3 + x$ and $u_2 = 1/3\theta x^3 - x$, where θ is a positive constant. Integrating $E_f u - E_g u$ by parts,

$$E_f u(x) - E_g u(x) = \int_0^1 u'(x)[G(x) - F(x)]\, dx \tag{18}$$

---

[10] Let w denote initial wealth and z, a random variable, be a risky prospect. The preference intensity function $v(w,z) = u[w+E(z)] - Eu(w+z)$ is the utility difference representing the strength of an individual's preference for the sure option $w+E(z)$ over the risk $w+z$. This is an alternative to the more commonly used risk premium, defined by $\pi(w,z) = w + E(z) - u^{-1}[Eu(w+z)]$. For a detailed account of the relationship between $v(w,z)$ and $\pi(w,z)$, see Bernell Stone.

[11] Alternatively, for risk preferrers, risk is transferred to those outcomes for which the individual's utility from risk is lower.

[12] Let $A(w) = -u''(x)/u'(x)$ be the absolute risk-aversion function. The sign of $\partial\pi(w,z)/\partial w$ is determined by the sign of $A'(w)$. Thus $u'''(w) > 0$ is a necessary but not sufficient condition for the risk premium to decrease as wealth increases. The *equivalence* between the sign of $u'''(w)$ and $\partial v(w,z)/\partial w$ has been shown by Hanson and Menezes.

so that

$$E_f u_1 - E_g u_1 = \int_0^1 (\theta x^2 + 1)[G(x) - F(x)]\,dx$$

$$E_f u_2 - E_g u_2 = \int_0^1 (\theta x^2 - 1)[G(x) - F(x)]\,dx$$

By our initial assumption,

$$\lim_{\theta \downarrow 0} [E_f u_1 - E_g u_1] = \int_0^1 [G(x) - F(x)]\,dx \geqslant 0$$

$$\lim_{\theta \downarrow 0} [E_f u_2 - E_g u_2] = -\int_0^1 [G(x) - F(x)]\,dx \geqslant 0$$

Thus $E_f u - E_g u \geqslant 0$ implies $\int_0^1 [G(x) - F(x)]\,dx = 0$, i.e., that $E_f = E_g$, which is (i).

To demonstrate (ii) consider another pair of utility functions in U_3^*: $u_3 = 1/3\theta x^3 + 1/2x^2$ and $u_4 = 1/3\theta x^3 - 1/2x^2$. Integrating (18) by parts and employing the fact that $E_f = E_g$,

$$(19) \quad E_f u - E_g u = -\int_0^1 u''(z) \int_0^z [G(x) - F(x)]\,dx\,dz$$

It follows that

$$\lim_{\theta \downarrow 0} [E_f u_3 - E_g u_3] = -\int_0^1 \int_0^z [G(x) - F(x)]\,dx\,dz \geqslant 0$$

$$\lim_{\theta \downarrow 0} [E_f u_4 - E_g u_4] = \int_0^1 \int_0^z [G(x) - F(x)]\,dx\,dz \geqslant 0$$

or $\int_0^1 \int_0^z [G(x) - F(x)]\,dx\,dz = 0$, which is (ii).

We prove (iii) by contradiction. Assume it is false. Then $\int_0^1 \int_0^y [G(z) - F(z)]\,dz\,dy < 0$ for some $x_0 \in [0, 1]$. By continuity, there exists an interval $[\alpha, \beta]$ containing x_0 such that, $\int_0^x \int_0^y [G(z) - F(z)]\,dz\,dy < 0$ for all $x \in [\alpha, \beta]$. When (19) is integrated by parts and simplified using (i) and (ii),

$$E_f u - E_g u = \int_0^1 u'''(x) \int_0^x \int_0^y [G(z) - F(z)]\,dz\,dy\,dx$$

For the following particular function in U_3^*,

$$u_5(x) = \begin{cases} 1/6x^3 & x \in [\alpha, \beta] \\ 1/6\theta x^3 & \text{otherwise} \end{cases}$$

$$\lim_{\theta \downarrow 0} [E_f u_5 - E_g u_5] = \int_0^x \int_0^y [G(z) - F(z)]\,dz\,dy\,dx < 0$$

which contradicts the original assumption of the theorem that $E_f u - E_g u \geqslant 0$ for all $u \in U_3^*$. Moreover, $\int_0^x \int_0^y [G(z) - F(z)]\,dz\,dy > 0$ for some $x \in (0, 1)$ since $G(x) \neq F(x)$.

The converse follows immediately by writing,

$$E_f u - E_g u = u'(b) \int_0^b [G(z) - F(z)]\,dz$$

$$- u''(b) \int_0^b \int_0^y [G(z) - F(z)]\,dz\,dy$$

$$+ \int_0^b u'''(x) \int_0^x \int_0^y [G(z) - F(z)]\,dz\,dy\,dx$$

Since $u'''(x) > 0$, (i), (ii), and (iii) of Theorem 1 guarantee that $E_f u \geqslant E_g u$ for all $u \in U_3^*$. This completes the proof.

It is clear from the discussion above, and the intuitive notion of downside risk, that both risk averters and risk preferrers can be downside risk averse. Similarly, individuals with increasing or decreasing utility functions can be averse to a leftward shift of risk. Thus $u'''(x) > 0$ is the *only* property of $u(x)$ necessary to define the set of downside risk averters. As indicated in the proof of the theorem, unanimous preference for $f(x)$ over $g(x)$ by such a diverse set of individuals requires equality of the mean and variance. In the last section we showed that the equality of the mean and variance is needed to ensure that an *MVPT* unambiguously transfers risk from the right to the left of any point. Here we have argued that this equality is important if we are not to arbitrarily exclude individuals who can be averse to a leftward shift of risk from the set of downside risk averters.[13]

Theorem 2 parallels the results obtained by Rothschild and Stiglitz who showed the equivalence of a definition of *increasing risk* based on mean-preserving spreads to one corresponding to unanimous preference by all individuals with concave utility functions. Let $U_2^* = \{u(x):\ u''(x) < 0\}$. They proved that a distribution $g(x)$ can be obtained from another $f(x)$ by a sequence of *MPSs* if and only if $f(x)$ is preferred to $g(x)$ by all individuals whose utility function belongs to U_2^*. The Rothschild-Stiglitz formulation requires equality for the means in order to analytically characterize the notion of increasing risk as a spreading of probability weight from the center to the tails of the distribution. Alternatively, since U_2^* includes *both* increasing and decreasing concave utility functions, unanimous preference for $f(x)$ over $g(x)$ requires that the means be the same.

IV. Comparison with the Literature

Two kinds of parametric formulations of downside risk have been widely used in the literature: probability loss and skewness measures. We now review these measures and relate them to our analysis. Later in this section we show how our general formulation of increasing downside risk is related to various stochastic dominance rankings of distributions.

Probability loss measures formulate downside risk as a probability-weighted function of deviations below an arbitrary target. The general form of these measures is

$$D_f(t,n) = \int_0^t (t-x)^n \, dF(x)$$

where t is the target and $n \geqslant 0$ determines the weights attached to deviations below the target. The probability of falling below the target, $D_f(t,0)$, is the basis of Roy's safety-first principle and has been suggested as a measure of downside risk by Markowitz and Pruitt. When $n=1$, D_f is the linearly weighted sum of probabilities of falling below the target used by Evsey Domar and Richard Musgrave. Target semivariance is obtained by setting $n=2$. It is the most widely used measure of downside risk.

There are two major problems with the probability loss measures of downside risk. For any particular individual with a specific target, each measure defines a complete ordering over the set of distributions. Since the orderings for individuals with different targets will not be the same, there may not be agreement about which of a pair of distributions has more downside risk. Thus the first basic problem is that, in general, any particular probability loss measure could only induce an ordering which corresponds to one based on unanimous agreement for a small and possibly very insignificant group of individuals. The second problem arises from the failure of the probability loss measures to distinguish increasing downside risk from other properties of distributions. For example, $g(x)$ may have more target semivariance than $f(x)$, i.e., $D_g(t,2) > D_f(t,2)$, because $g(x)$ has either a smaller mean, larger variance, or more dispersion below the target than $f(x)$.

The only probability loss measure that is directly related to our definition of downside risk is target semivariance. The relationship follows directly from the analysis of Section II. Specifically,

PROPOSITION 1:

A. *If $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs, then $g(x)$ has at least as much semivariance as $f(x)$ for any target $t \in [0,1]$.*

---

[13] It is interesting to note that Archibald Nichol's early paper on firm behavior under uncertainty reaches similar conclusions. He defines a conservative management as one "... more influenced by chances of loss and chances of net revenue less than the maximum mean of expectations...." and a speculative management as one "... more influenced by chances of net revenue greater than the maximum mean of expectations" (p. 652). He then argues that while the impact of an increase in risk (dispersion) depends on whether the firm is conservative or speculative, both types of managers will react in the same way to changes in skewness.

B. *Let $g(x)$ and $f(x)$ be distributions with the same mean and variance. If the semivariance of $g(x)$ is at least as large as that of $f(x)$ for each target $t \in [0,1]$ and strictly larger for t in some subinterval, then $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs.*

Thus, as might be expected, for target semivariance to order a pair of distributions in terms of downside risk, the distributions must have the same mean and variance and in addition the semivariance of one distribution must be at least as large as that of the other for *all* targets and strictly larger for targets in some interval.

Skewness, as measured by the third central moment of a distribution, is another widely used measure of downside risk.[14] The more right skewed a distribution, the greater is the probability-weighted sum of cubed deviations above the mean relative to those below it. Given equality of the means and variances, distributions which are more skewed to the right are regarded as providing better downside protection or smaller downside risk. The disadvantages of skewness as a measure of downside risk are analogous to those of variance as a measure of risk. It is easy to show that $f(x)$ may be preferred to $g(x)$ by some downside risk-averse individuals even though $f(x)$ and $g(x)$ have the same mean and variance and $g(x)$ is more skewed to the right.

The relationship between our formulation of increasing downside risk and skewness is summarized in the following propositions.

PROPOSITION 2:

A. *If $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs, then $f(x)$ is more right skewed than $g(x)$ (i.e., $S_f > S_g$).*

B. *Let $g(x)$ and $f(x)$ be distributions with the same mean and variance. Then $f(x)$ more skewed to the right than $g(x)$ does not imply that $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs.*

PROOF:

We first prove Proposition 2A. Let $g(x) = f(x) + \Sigma_i t_i(x)$, where $t_i(x)$ is an *MVPT*. By definition,

$$S_g = \int_0^1 (x-\mu)^3 f(x)\,dx + \sum_i \int_0^1 (x-\mu)^3 t_i(x)\,dx = S_f + \sum_i \int_0^1 (x-\mu)^3 t_i(x)\,dx$$

Expanding $(x-\mu)^3$ in the second term and recalling that

$$\int_0^1 x^2 t(x)\,dx = \int_0^1 x t(x)\,dx = \int_0^1 t(x) = 0$$

we obtain $S_g = S_f + \sum_i \int_0^1 x^3 t_i(x)\,dx$

Condition (12) guarantees that

$$\int_0^1 \int_0^x T^*(y)\,dy\,dx > 0$$

Integrating by parts, using the relevant properties of T^* and T, and summing gives

$$\sum_i \int_0^1 \int_0^x T_i^*(y)\,dy\,dx = -1/6 \sum_i \int_0^1 x^3 t_i(x)\,dx > 0$$

so that $S_f > S_g$. The validity of Proposition 2B follows immediately from the fact that $E_f = E_g$, $V_f = V_g$ and $S_f > S_g$ do not guarantee that (12) holds.[15]

We have shown that if a distribution has more downside risk than another, it is less

---

[14] See the references cited in fn. 2.

[15] The following pair of risks provide an example of the validity of Proposition 2B:

$X=x$	0	1	2	3
$f(x)$	0.1	0.5	0.2	0.2
$g(x)$	0.2	0.2	0.5	0.1

This risk pair has been used by Whitmore (1975) in discussing the relation between skewness and third-degree stochastic dominance.

skewed to the right. However, the converse is not necessarily true.

We now show how our analysis is related to the stochastic dominance (*SD*) literature. The *SD* approach begins by identifying a meaningful group of individuals whose utility functions belong to a well-defined class U. The problem is then to define the same partial order on the set of distributions as is induced by the class U, but to use only characteristics of the distributions in defining it. Our approach, like that of Rothschild and Stiglitz, falls within the *SD* framework. By defining increasing downside risk in terms of *MVPT*s we get the same partial order as is obtained by unanimous agreement on the part of all downside risk averters, that is, those whose utility functions belong to U_3^*.

The three most widely used *SD* formulations are first, second, and third-degree stochastic dominance, or *FSD*, *SSD*, and *TSD*, respectively.[16] These are the orderings resulting from unanimous preference over the respective classes,

$$U_1=\{u(x):u'(x)>0\}$$
$$U_2=\{u(x)\in U_1:u''(x)<0\}$$
$$U_3=\{u(x)\in U_2:u'''(x)>0\}$$

The corresponding *SD* rules are

FSD: $E_f u(x)\geqslant E_g u(x)$ for all $u(x)$ in U_1 iff $F(x)\neq G(x)$ and $F(x)\leqslant G(x)$ for all $x\in[0,1]$

SSD: $E_f u(x)\geqslant E_g u(x)$ for all $u(x)$ in U_2 iff $F(x)\neq G(x)$ and $\int_0^x F(y)\,dy\leqslant\int_0^x G(y)\,dy$ for all $x\in[0,1]$

TSD: $E_f u(x)\geqslant E_g u(x)$ for all $u(x)$ in U_3 iff $F(x)\neq G(x)$, $E_f\geqslant E_g$ and $\int_0^x\int_0^y F(z)\,dz\,dy\leqslant\int_0^x\int_0^y G(z)\,dz\,dy$ for all $x\in[0,1]$.

By successively narrowing the class of utility functions under consideration, $U_3\subset U_2\subset U_1$, these *SD* rules increase the number of pairs of distributions which can be compared. In contrast, our analysis and that of Rothschild and Stiglitz identify the classes of utility functions U_3^* and U_2^* required to define increasing downside risk and increasing risk in a choice-theoretic setting. It follows immediately from the definition of U_3^* and U_2^* that neither set is a subset of the other, that is, a risk averter may or may not be downside risk averse, and vice versa. In terms of the *SD* rules, if $f(x)$ dominates $g(x)$ by *TSD*, then $f(x)$ also dominates $g(x)$ by *SSD* and *FSD*. However, if $f(x)$ and $g(x)$ are comparable by increasing downside risk, they are noncomparable by increasing risk.

We conclude our survey of the literature by showing the precise relationship between our analysis and *SD*. Since $U_3\subset U_3^*$ but U_3^* is noncomparable with U_2 or U_1, the only conventional *SD* criterion that is directly related to our definition of increasing downside risk is *TSD*. The propositions below formally state the relationship.

PROPOSITION 3:

A. *If $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs, then $f(x)$ dominates $g(x)$ by TSD.*

B. *Let $g(x)$ and $f(x)$ be distributions with the same mean and variance. If $f(x)$ dominates $g(x)$ by TSD then $g(x)$ can be obtained from $f(x)$ by a sequence of MVPTs.*

These relationships follow immediately from a comparison of the *TSD* rule with conditions (i), (ii), and (iii) of Theorem 1.

V. Summary and Conclusions

In this paper three definitions of increasing downside risk are formulated and shown to be equivalent. The first is based on spread-contraction combinations which unambiguously transfer risk leftward in a distribution while maintaining the mean and variance. This definition analytically captures the intuitive and empirically relevant notions of downside risk and distinguishes it from other characteristics of distributions, for example, riskiness. To provide a choice-theoretic foundation for our analysis we defined increasing downside risk in terms of the unanimous choice of all individuals with convex marginal utility functions, that is, all individuals who are decreasingly risk averse. One implication of this result is that both risk averters and risk preferrers can be downside risk averse. The third definition provides a practical criterion to compare distributions in terms of downside risk. We have also shown how our general approach is related to the prevailing measures of downside risk on the one hand, and stochastic dominance on the other.

---

[16] See James Quirk and Rubin Saposnik, Josef Hadar and William Russell, Hanoch and Levy, and Whitmore (1970).

REFERENCES

C. P. Alderfer and H. Bierman, "Choices with Risk: Beyond the Mean and Variance," *J. Bus., Univ. Chicago*, July 1970, *43*, 341–53.

F. D. Arditti, "Risk and the Required Return on Equity," *J. Finance*, Mar. 1967, *22*, 19–36.

W. J. Baumol, "An Expected Gain-Confidence Limit Criterion for Portfolio Selection," *Manage. Sci.*, Oct. 1963, *10*, 174–82.

A. Charnes and W. W. Cooper, "Chance-Constrained Programming," *Manage. Sci.*, Oct. 1959, *6*, 73–79.

C. H. Coombs and D. G. Pruitt, "Components of Risk in Decision-Making: Probability and Variance Preferences," *J. Experim. Psychol.*, Nov. 1960, *60*, 265–77.

E. V. Domar and R. A. Musgrave, "Proportional Income Taxation and Risk Taking," *Quart. J. Econ.*, May 1944, *58*, 384–422.

P. C. Fishburn, "Mean-Risk Analysis with Risks Associated with Below-Target Returns," *Amer. Econ. Rev.*, Mar. 1977, *67*, 116–24.

M. Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *J. Polit. Econ.*, Aug. 1948, *56*, 279–304.

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

F. H. Hahn, "Savings and Uncertainty," *Rev. Econ. Stud.*, Jan. 1970, *37*, 21–24.

G. Hanoch and H. Levy, "Efficient Portfolio Selection with Quadratic and Cubic Utility," *J. Bus., Univ. Chicago*, Apr. 1970, *43*, 181–89.

D. L. Hanson and C. F. Menezes, "On a Neglected Aspect of the Theory of Risk Aversion," *Western Econ. J.*, June 1971, *9*, 211–17.

John R. Hicks, *Critical Essays in Monetary Theory*, Oxford 1967.

R. F. Lanzillotti, "Pricing Objectives in Large Companies," *Amer. Econ. Rev.*, Dec. 1958, *48*, 921–40.

R. Libby and P. C. Fishburn, "Behavioral Models of Risk Taking in Business Decisions: A Survey and Evaluation," *J. Account. Res.*, Autumn 1977, *15*, 272–92.

J. C. T. Mao, (1970a) "Models of Capital Budgeting, *E-V* vs *E-S*," *J. Finan. Quant. Anal.*, Jan. 1970, *4*, 657–75.

______, (1970b) "Survey of Capital Budgeting: Theory and Practice," *J. Finance Proc.*, May 1970, *25*, 349–60.

Harry M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, New Haven 1959.

R. T. Masson, "Utility Functions with Jump Discontinuities: Some Evidence and Implications from Peasant Agriculture," *Econ. Inquiry*, Dec. 1974, *12*, 559–66.

A. J. Nichol, "Probability Analysis in the Theory of Demand, Net Revenue and Price," *J. Polit. Econ.*, Oct. 1941, *49*, 637–61.

D. G. Pruitt, "Pattern and Level of Risk in Gambling Decisions," *Psychol. Rev.*, May 1962, *69*, 187–201.

J. P. Quirk and R. Saposnik, "Admissibility and Measurable Utility Functions," *Rev. Econ. Stud.*, Feb. 1962, *29*, 140–46.

R. N. Rosett, "Estimating the Utility of Wealth from Call Option Data," in Donald D. Hester and James Tobin, eds., *Risk and Portfolio Choice*, Cowles Found. Monog. No. 19, New York 1967.

M. Rothschild and J. Stiglitz, "Increasing Risk: I. A Definition," *J. Econ. Theory*, Sept. 1970, *2*, 225–43.

A. D. Roy, "Safety First and the Holding of Assets," *Econometrica*, July 1952, *20*, 431–49.

D. G. Simonson, "The Speculative Behavior of Mutual Funds," *J. Finance Proc.*, May 1972, *28*, 381–91.

Bernell K. Stone, *Risk, Return and Equilibrium*, Cambridge Mass., 1970.

S. C. Tsiang, "The Rationale of the Mean-Standard Deviation Analysis, Skewness

Preference, and the Demand for Money," *Amer. Econ. Rev.*, June 1972, *62*, 354–71.

R. N. Waud, "Asymmetric Policymaker Utility Functions and Optimal Policy Under Uncertainty," *Econometrica*, Jan. 1976, *44*, 53–66.

G. A. Whitmore, "The Theory of Skewness Preference," *J. Bus. Admin.*, Fall 1975, *6*, 13–20.

______, "Third Degree Stochastic Dominance," *Amer. Econ. Rev.*, June 1970, *60*, 457–59.

Compensating Variation, Consumer's Surplus, and Welfare

By John S. Chipman and James C. Moore[*]

The compensating variation was defined by J. R. Hicks (1942) as the amount one would have to deduct from a person's income to make him just as well off after a change in prices and income as he had been in the initial situation. If the compensating variation is positive, the individual is better off under the new situation. Since the compensating variation furnishes a uniquely defined numerical (cardinal) indicator of welfare improvement, it provides an implicit ranking of alternative prospective situations not only relative to the initial situation, but also relative to each other. In practice this appears to be how this and other tools of cost-benefit analysis are actually used: one is interested in knowing not only whether a particular bridge, or a particular excise tax, will lead to an improvement in welfare, but which out of a set of alternative bridges or alternative tax systems will improve welfare the most.[1] In this paper we analyze conditions under which the compensating variation can be validly used in this generalized sense; these turn out to be precisely the same as conditions previously derived (see our 1976 paper) for the valid use of consumer's surplus as a welfare measure.

In our final section we analyze the problem of deriving the generalized equivalent and compensating variations (either exactly or approximately) from observable demand functions, by means of generalizations of, or alternatives to, consumer's surplus.

I. Consumer's Surplus and the Compensating Variation

We assume that the consumer has a preference ordering R defined on the commodity space (which we take to be the n-dimensional nonnegative orthant), with consumption bundles denoted $x=(x_1, x_2, \ldots, x_n)$. The consumer is assumed to have a single-valued, continuously differentiable demand function $h(p, m)=(h^1(p, m), h^2(p, m), \ldots, h^n(p, m))$, where h^j is the demand for commodity j, $p=(p_1, p_2, \ldots, p_n)$ is a vector of strictly positive prices, and m (also strictly positive) denotes the consumer's income. The demand function is assumed to be generated by R, so that $h(p,m)Rx$ for any x satisfying the budget constraint $p\cdot x \leqslant m$; and we assume that all income is spent, i.e., $p\cdot h(p,m)=m$ for all (p,m). Pairs (p,m) will be called *budgets*, and a particular set of budgets under consideration will be called the *budget domain*. The *indirect preference relation* R^* induced by R is defined by the condition

$$(1) \qquad (p^1, m^1)R^*(p^2, m^2) \text{ iff } h(p^1, m^1)Rh(p^2, m^2)$$

Owing to the homogeneity of degree zero of the demand function (absence of money illusion), the relation R^* is also "homogeneous of degree zero" in the sense that $(p^1, m^1)R^*(p^2, m^2)$ implies $\lambda_1(p^1, m^1) R^*\lambda_2(p^2, m^2)$ for any $\lambda_1, \lambda_2>0$. We shall denote by I^* the corresponding indirect indifference relation, defined by $(p^1, m^1) I^*(p^2, m^2)$ iff $(p^1, m^1)R^*(p^2, m^2)$ and $(p^2, m^2)R^*(p^1, m^1)$.

Fundamental to our approach is the principle that the consumer is the best judge of his own welfare. As a basic principle of social policy, this can of course be questioned, certainly in particular cases (such as that of chemical addiction); we merely wish to emphasize that we accept this principle as our basic working hypothesis in this paper. A welfare indicator such as consumer's surplus or the compensating variation must, according to this principle, be judged by the criterion of whether or not it correctly represents the consumer's indirect preferences.

---

[*] University of Minnesota and Purdue University, respectively. Research for this paper was supported by National Science Foundation grant SOC77-27257.

[1] As Arnold Harberger has stated:

> What could be more useful to us as a guide to priorities in tax reform than the knowledge that the deadweight losses stemming from the tax loopholes (percentage depletion and capital gains) open to explorers for oil and gas are probably greater in total magnitude than the deadweight losses associated with all the other inefficiencies induced by the corporation income tax?
>
> [1964, p. 58–59]

See also the cogent remarks by John Hause, p. 1150.

Just as a (direct) utility function U defined on the commodity space is said to represent the (direct) preference relation R if $U(x^1) \geqslant U(x^2)$ iff $x^1 R x^2$, and *indirect utility function* V defined on the relevant budget space is a numerical function which represents the indirect preference relation R^* in the sense that

$$(2) \quad V(p^1, m^1) \geqslant V(p^2, m^2) \quad \text{iff } (p^1, m^1) R^* (p^2, m^2)$$

Such a function V will also be termed a *numerical indicator of indirect preferences.* A welfare indicator will be described as an *acceptable measure of welfare change* if it is a numerical indicator of change in the consumer's indirect preferences (possibly on a restricted budget domain), hence if it is an increment in an indirect utility function (at least when restricted to this domain).

Let $v = (v^1, v^2, \ldots, v^{n+1})$ be an $(n+1)$-vector-valued continuously differentiable function of (p, m), and let ω be a path in the budget space, that is, a continuous function $\omega(t)$, $0 \leqslant t \leqslant 1$, with endpoints $\omega(0) = (p^0, m^0)$ and $\omega(1) = (p^1, m^1)$, say. The line integral of v with respect to ω may be denoted

$$(3) \quad \int_0^1 v[\omega(t)] \cdot d\omega(t) = \sum_{j=1}^{n+1} \int_0^1 v^j[\omega(t)]\, d\omega_j(t)$$

It is well-known from the theory of line integrals (see Tom Apostol, pp. 276–93) that if all polygonal paths ω joining two budgets (p^0, m^0) and (p^1, m^1) yield the same value for the line integral (3) (i.e., if the line integral is *path independent*), then there exists a twice-differentiable "potential function" $V(p, m)$ such that

$$(4) \quad \int_0^1 v[\omega(t)] \cdot d\omega(t) = V(p^1, m^1) - V(p^0, m^0)$$

and satisfying

$$(5) \quad \frac{\partial V(p,m)}{\partial p_j} = v^j(p,m) \qquad (j = 1, 2, \ldots, n);$$

$$\frac{\partial V(p,m)}{\partial m} = v^{n+1}(p,m)$$

Conversely, if there exists a twice-differentiable function $V(p, m)$ satisfying (5), then the line integral (3) is path independent, and, in fact, (4) is satisfied.

We shall say that a continuously differentiable $(n+1)$-vector-valued function $v(p, m)$ *furnishes an acceptable integral measure of welfare change* iff: (i) the line integral (3) is path independent, and (ii) it is nonnegative iff $(p^1, m^1) R^* (p^0, m^0)$, i.e., iff the terminal point is indirectly preferred or indifferent to the initial point. In terms of the above terminology, therefore, the line integral (3) satisfying these properties is an acceptable measure of welfare change.

If $V(p, m)$ is a differentiable indirect utility function representing R^*, then Antonelli's partial differential equations are satisfied:[2]

$$(6) \quad \frac{\partial V(p,m)}{\partial p_j} = -\frac{\partial V(p,m)}{\partial m} h^j(p,m) \qquad \text{for } j = 1, 2, \ldots, n$$

---

[2] See Antonelli (1886, p. 17; 1971, p. 349, equation (24) and the unnumbered equation following it). In the recent literature these equations (which are no more "identities" than are any other partial differential equations) have been described as (as well as mispronounced) "Roy's identity," even though René Roy's was a second rediscovery, the first being that of R. G. D. Allen, p. 190. See Chipman et al., pp. 348-49, fn. 51; see also our 1976 paper, pp. 73–74, where a simple proof is presented not requiring differentiability of the direct utility function U.

If, then, we choose for v the function

$$(7)\qquad v^j(p,m) = -\frac{\partial V(p,m)}{\partial m}h^j(p,m) \qquad (j=1,2,\ldots,n)$$

$$v^{n+1}(p,m)=\frac{\partial V(p,m)}{\partial m}$$

it is clear that v furnishes an acceptable integral measure of welfare change.

If in place of the function v of (7) we choose the function $f=(f^1,f^2,\ldots,f^{n+1})$ defined by

$$(8)\qquad f^j(p,m)=-h^j(p,m) \qquad (j=1,2,\ldots,n)$$

$$f^{n+1}(p,m)=1$$

the corresponding line integral (where $\omega(t)=(p(t),m(t))$)

$$(9)\qquad \int_0^1 f[\omega(t)]\cdot d\omega(t) = -\sum_{j=1}^{n}\int_0^1 h^j[\omega(t)]\,dp(t)+m(1)-m(0)$$

is the *Marshallian consumer's surplus.*

Let us now turn to the Hicksian compensating variation and Hicks' alternative definition of consumer's surplus. This is best analyzed by means of the Hurwicz and Uzawa *income-compensation function,* defined by[3]

$$(10)\qquad \mu(p;p^0,m^0) = \min\{m>0:(p,m)R^*(p^0,m^0)\}$$

It has the following properties:

M1 $\mu(p;\lambda p^0,\lambda m^0)=\mu(p;p^0,m^0)$ for $\lambda>0$

M2 $\mu(\lambda p;p^0,m^0)=\lambda\mu(p;p^0,m^0)$ for $\lambda>0$

M3 $\mu(p^0;p^0,m^0)=m^0$

M4 $(p^0,m^0)I^*(p,m)$ iff $m^0=\mu(p^0;p,m)$

Most noteworthy is the property that (for continuous R^*) the function V_{p^t} defined for any fixed p^t by

$$(11)\qquad V_{p^t}(p,m)=\mu(p^t;p,m)$$

is an indirect utility function representing R^*.[4]

Given a base budget (p^0,m^0), we can use the income-compensation function to define the *compensated demand function* by

$$(12)\qquad g(p;p^0,m^0)=h[p,\mu(p;p^0,m^0)]$$

It was shown by Hurwicz and Uzawa (p. 120) that

$$(13)\qquad \mu_j(p;p^0,m^0)=g^j(p;p^0,m^0) \qquad (j=1,2,\ldots,n)$$

where μ_j denotes the partial derivative of μ with respect to its jth argument, and g^j is the compensated demand for commodity j.

Hicks (1942) defined the *equivalent variation* in going from (p^0,m^0) to (p^1,m^0) by

$$(14)\qquad E(p^0;p^1,m^0)=\mu(p^0;p^1,m^0)-m^0$$

The natural welfare indicator which extends this definition to comparison between any two budgets (p^1,m^1) and (p^2,m^2) is (for fixed base budget (p^0,m^0))

$$(15)\qquad \Phi(p,m)=\mu(p^0;p,m)-m^0$$

We may call (15) the *generalized equivalent variation.* Since it differs from V_{p^0} by a constant (see (11)), it is clear that it is an indirect utility function representing the indirect preference relation R^*, hence is an acceptable measure of welfare change (see Hause).

---

[3] Based on Lionel McKenzie's "minimum-income function" $M_x(p)$ to which it is related by setting $x=h(p^0,m^0)$.

[4] For a proof, see Lemma 8 of Leonid Hurwicz and Hirofumi Uzawa, pp. 129–30.

Hicks (1942) defined the *compensating variation* in going from (p^0, m^0) to (p^1, m^0) by

$$C(p^1; p^0, m^0) = m^0 - \mu(p^1; p^0, m^0) \tag{16}$$

The natural welfare indicator which extends this definition to comparison between any two budgets (p^1, m^1) and (p^2, m^2) is (for fixed (p^0, m^0))

$$\Psi(p, m) = m - \mu(p; p^0, m^0) \tag{17}$$

since it yields $\Psi(p^1, m^0) - \Psi(p^0, m^0) = C(p^1; p^0, m^0)$. We shall therefore call (17) the *generalized compensating variation*. Unlike the generalized equivalent variation, however, the generalized compensating variation is not a proper numerical indicator of indirect preferences except for a limited set of comparisons or under restricted conditions. (This point has been made by Hause.) For comparisons between an arbitrary budget (p, m) and the given base budget (p^0, m^0), we have from (17) and property M3,

$$\begin{aligned}\Psi(p, m) - \Psi(p^0, m^0) \\ = \mu(p; p, m) - \mu(p; p^0, m^0)\end{aligned} \tag{18}$$

hence the change in the generalized compensating variation coincides with the increment in the Hurwicz-Uzawa function V_p of (11); but Ψ cannot be used to represent indirect preferences as between two budgets (p^1, m^1), (p^2, m^2) both distinct from (p^0, m^0), except under limited conditions, two of which we take up in Sections II and III below.[5]

Consider now the vector-valued function $\psi = (\psi^1, \psi^2, \dots, \psi^{n+1})$, defined for fixed (p^0, m^0) by

$$\begin{aligned}\psi^j(p, m) &= -g^j(p; p^0, m^0) \\ &\qquad (j = 1, 2, \dots, n) \\ \psi^{n+1}(p, m) &= 1\end{aligned} \tag{19}$$

Using this function ψ in place of the function f of (8) we obtain for the line integral of ψ with respect to ω

$$\begin{aligned}\textstyle\int_0^1 \psi[\omega(t)] \cdot d\omega(t) \\ = -\sum_{j=1}^{n} \textstyle\int_0^1 g^j(p(t); p^0, m^0)\, dp(t) \\ + m(1) - m(0)\end{aligned} \tag{20}$$

This is the *Hicksian consumer's surplus*. Since from (17), (13), and (19) we see that $\partial\Psi/\partial p_j = \psi^j$ for $j = 1, 2, \dots, n$, and $\partial\Psi/\partial m = \psi^{n+1}$, it follows that the line integral of ψ with respect to ω, connecting the endpoints $\omega(0) = (p^1, m^1)$ and $\omega(1) = (p^2, m^2)$, is precisely

$$\textstyle\int_0^1 \psi[\omega(t)] \cdot d\omega(t) = \Psi(p^2, m^2) - \Psi(p^1, m^1) \tag{21}$$

so that Ψ is a potential function for this line integral. For $\omega(0) = (p^0, m^0)$ and $\omega(1) = (p^1, m^0)$, the Hicksian consumer's surplus is then precisely Hicks' compensating variation (16).

We are now ready to formulate our problem precisely: under what conditions (on preferences and on admissible pairs of endpoints (p^1, m^1) and (p^2, m^2) under comparison) do the Marshallian and Hicksian consumer's surplus integrals (9) and (20) furnish acceptable integral measures of welfare change? The second of these questions is equivalent, as we have just seen, to the question: under what conditions does the generalized compensating variation (17) constitute an acceptable measure of welfare change?

In our 1976 paper, three answers were provided to the above question concerning the Marshallian consumer's surplus integral:

---

[5] The problem is that, if we wish to use (17) to compare two alternative budgets (p^1, m^1) and (p^2, m^2), then it follows from (18) and (11) that $\Psi(p^2, m^2) > \Psi(p^1, m^1)$ only if the increment in V_{p2} is at least as great as the increment in V_{p1}. While V_{p1} and V_{p2} are both perfectly valid indirect utility functions, there is no reason to expect a change according to one such representation to be comparable to a change according to another.

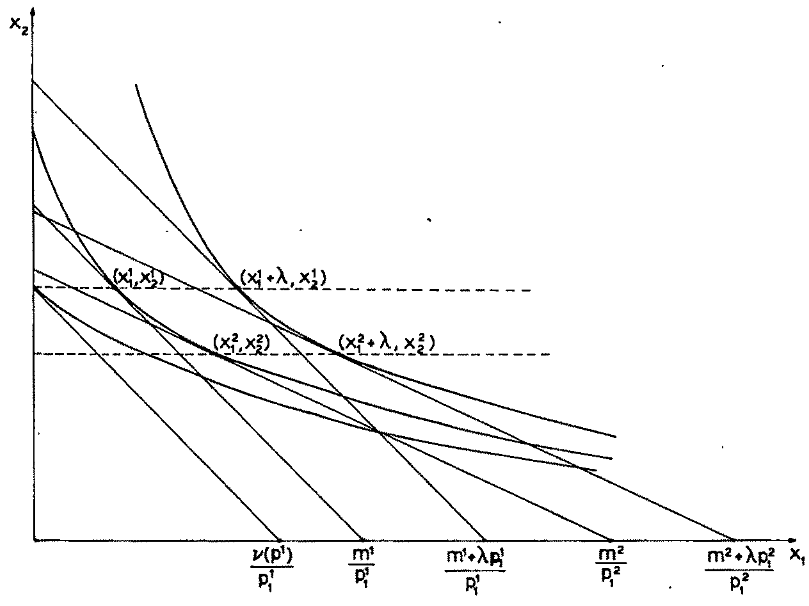


FIGURE 1

(a) If there is no restriction on the pairs of endpoints (p^1, m^1), (p^2, m^2) under comparison, then there are no conditions on preferences under which the consumer's surplus integral (9) furnishes an acceptable measure of welfare change.

(b) If the pairs of endpoints (p^1, m^1) and (p^2, m^2) that may be compared are restricted by the condition that $m^1 = m^2$, but are otherwise free to vary arbitrarily, then the Marshallian consumer's surplus furnishes an acceptable measure of welfare change *if and only if preferences are homothetic.*[6]

(c) If the pairs of endpoints (p^1, m^1) and (p^2, m^2) under comparison are constrained by the condition that $p_1^1 = p_1^2$ (or equivalently, if income and all other prices are deflated by the price of commodity 1 in each period), but are otherwise free to vary arbitrarily within a certain budget domain,[7] then the Marshallian consumer's surplus integral furnishes an acceptable measure of welfare change *if and only if preferences are parallel with respect to commodity 1 on the relevant domain of the commodity space* (see Figure 1).[8]

Turning now to the generalized compensating variation (17), we note that (using property M2) it has the property of being homogeneous of degree one in (p, m). On the other hand, for it to serve as an indirect utility function it would have to be homogeneous of degree zero in (p, m), as well as strictly increasing in m (owing to the fact that h is assumed to satisfy the budget equality $p \cdot h(p, m) = m$). These properties are manifestly incompatible. Our answer (a) above therefore applies just as well to the generalized compensating variation and the Hicksian consumer's surplus as to the Marshallian consumer's surplus. In the following two sections we show that the same is true for (b) and (c); that is, the necessary and sufficient conditions for the Hicksian consumer's surplus to furnish an acceptable measure of welfare change are precisely the same as those for the Marshallian consumer's surplus to do so.[9] In the concluding section we analyze the relation between the Marshallian and Hicksian measures, and take up the question of the feasibility of estimating the Hurwicz-Uzawa function (11) from empirically estimated demand functions.

---

[6] A preference relation R is, as usual, termed *homothetic* if $x^1 R x^2$ iff $\lambda x^1 R \lambda x^2$ for $\lambda > 0$.

[7] The domain in question is defined by the requirement that m be greater than some minimum income $\nu(p)$. See equation (35) below.

[8] A preference relation R will be called *parallel with respect to commodity* 1 if commodity 1 is a strictly preferred commodity (in the sense that whenever $x_1^1 > x_1^2$ and $x_i^1 = x_i^2$ for $i = 2, 3, \ldots, n$, x^1 is strictly preferred to x^2) and moreover, $x^1 R x^2$ iff $x^1 + \lambda e^1 R x^2 + \lambda e^1$ for $\lambda > \max\{-x_1^1, -x_1^2\}$, where $e^1 = (1, 0, 0, \ldots, 0)$. This is equivalent to saying that if a consumer is indifferent between two bundles x^1 and x^2, he will remain indifferent between them if the same amount λ of commodity 1 is added to (or subtracted from) each bundle (see Figure 1). An example of parallel preferences in the two-commodity case is furnished by the utility function $U(x_1, x_2) = x_1 + \log x_2$. More generally, parallel preferences can be represented by utility functions of the form $U(x) = a x_1 + W(x_2, x_3, \ldots, x_n)$; see F. Y. Edgeworth, p. 237n, Paul Samuelson (1942, p. 85), and our 1976 paper, p. 88. While a formal treatment of the concept of parallel preferences does not appear to have been undertaken prior to our 1976 paper, the concept goes back to Kenneth Boulding and even further to Edgeworth, and the need for the explicit assumption of parallel preferences in consumer's surplus analysis has been emphasized by Samuelson (1964) and recognized by a number of other writers, for example, Edwin Mansfield, pp. 68–69, and Hicks himself (1939, p. 40).

II. The Case of Fixed Income and Homothetic Preferences

We take up first the case in which income remains unchanged.[10] We show that for the generalized compensating variation to be an acceptable measure of welfare change in this case, it is necessary and sufficient that preferences be homothetic.

Let $\bar{m} > 0$ be the fixed value of the consumer's income. Then the generalized compensating variation (17) correctly represents the consumer's indirect preference relation R^* on a budget domain with $(p, m) = (p, \bar{m})$ iff for a given base budget (p^0, m^0),

$$(22) \quad (p^1, \bar{m}) R^* (p^2, \bar{m}) \quad \text{iff } \bar{m} - \mu(p^1; p^0, m^0) \geqslant \bar{m} - \mu(p^2; p^0, m^0)$$

Our demonstration will be set out in the form of a series of seven propositions:

PROPOSITION H1: *Condition* (22) *is satisfied if and only if indirect preferences* R^* *are representable by the indirect utility function*

$$(23) \quad V(p, m) = \frac{m}{\mu(p; p^0, m^0)}$$

on the homogeneously extended budget domain.[11]

PROOF:

If (p^1, m^1) and (p^2, m^2) are any two budgets such that $(p^1, m^1) R^* (p^2, m^2)$, by the homogeneity of degree 0 of R^* we have $(\bar{m} p^1/m^1, \bar{m}) R^* (\bar{m} p^2/m^2, \bar{m})$. Now, applying (22) and using property M2 of μ, we obtain $m^1/\mu(p^1; p^0, m^0) \geqslant m^2/\mu(p^2; p^0, m^0)$. Conversely, by reversing the steps in the argument we see that the latter inequality implies $(p^1, m^1) R^* (p^2, m^2)$. Thus,

$$(p^1, m^1) R^* (p^2, m^2) \quad \text{iff} \frac{m^1}{\mu(p^1; p^0, m^0)} \geqslant \frac{m^2}{\mu(p^2; p^0, m^0)}$$

PROPOSITION H2: *If* R^* *is representable by* (23), *the demand function satisfies*

$$(24) \quad h(p, \lambda m) = \lambda h(p, m) \qquad \text{for } \lambda > 0$$

hence the direct preference relation R *is homothetic.*

This follows by applying the Antonelli equations (6) to (23), to obtain

$$(25) \quad h^j(p, m) = m \frac{\mu_j(p; p^0, m^0)}{\mu(p; p^0, m^0)} = m h^j(p, 1)$$

---

[9] It is no accident that the Hicksian and Marshallian measures are related in this way in the case of the two types of restrictions on the domain we have considered here, namely constant income and constant price of the numeraire. We plan to show in a future paper that this relationship holds for any differentiable restriction on the budget domain.

[10] Harberger's (1971, p. 788) specific exclusion of income increments (which he calls the "first-order change in utility") from his consumer's surplus concept suggests that this case may be relevent to his work. Furthermore, he assumes some so-called "integrability conditions" (1964, p. 62) that are known (when applied to demand functions) to imply homotheticity of preferences (see our 1976 paper p. 95, Lemma 6; also Samuelson, 1942, p. 81n; and Eugene Silberberg, p. 946).

[11] By "homogeneously extended budget domain" we mean the following: If the original budget domain consists of budgets $(\pi, \bar{m})$ with π in a set Π, then the homogeneously extended budget domain is the set of budgets $(p, m) = \lambda(\pi, \bar{m})$ with $\lambda > 0$ and π in Π, or equivalently, the set of budgets (p, m) such that $(\bar{m}/m)p$ is in Π. See our 1979 paper.

where the last equality follows by setting $m=1$ in the first. Thus, $h^j(p, \lambda m) = \lambda m h^j(p,1) = \lambda h^j(p, m)$.

It is well known that (24) is necessary and sufficient for the preference relation R to be homothetic (see Chipman).

Propositions H1 and H2 together imply that homotheticity of preferences is a necessary condition for the compensating variation to be an acceptable measure of welfare change when income is constant and prices are free to vary arbitrarily. The sufficiency follows from the remaining five propositions.

PROPOSITION H3: *If the consumer's preference relation R is homothetic, then the indirect preference relation R^* satisfies*

$$(26)\quad (p^1, m^1)R^*(p^2, m^2)$$

$$\textit{iff}\ (p^1, \lambda m^1)R^*(p^2, \lambda m^2) \qquad \textit{for}\ \lambda > 0$$

Since (as noted above) homotheticity implies (24), the result follows from homotheticity, (24), and application of the definition (1) of R^*.

PROPOSITION H4: *If an indirect preference relation R^* satisfies* (26) *then, for any fixed price vector p^t the Hurwicz and Uzawa function* (11) *has the property*

$$(27)\quad V_{p^t}(p, \lambda m) = \lambda V_{p^t}(p, m) \qquad \textit{for}\ \lambda > 0$$

By M4 we have $(p^t, \mu(p^t; p, m))I^*(p, m)$, whence it follows from Proposition H3 that $(p^t, \lambda\mu(p^t; p, m))I^*(p, \lambda m)$. Applying M4 once again we obtain $\mu(p^t; p, \lambda m) = \lambda\mu(p^t; p, m)$.

PROPOSITION H5: *Property* (27) *implies that given any two base price vectors p^1 and p^2, the corresponding Hurwicz and Uzawa indirect utility representations V_{p^t} $(t=1,2)$ are related by*

$$(28)\quad V_{p^1}(p, m) = V_{p^1}(p^2, 1)V_{p^2}(p, m)$$

for all budgets (p, m), where $V_{p1}(p^2, 1) > 0$; that is, given any two such representations for R^*, one is a constant positive multiple of the other.

From the definition (11) of V_{pt} and by property M4 of μ, we have by (27):

$$\mu(p^1; p, m) = \mu(p^1; p^2, \mu(p^2; p, m))$$

$$= \mu(p^2; p, m)\mu(p^1; p^2, 1)$$

where $\mu(p^1; p^2, 1) > 0$ by the positive homogeneity of $\mu(p^1; p, m)$ in m and the fact that it is strictly increasing in m.

PROPOSITION H6: *Condition* (28) *implies that μ satisfies the functional equation*[12]

$$(29)\quad \frac{\mu(p^1; p', m')}{\mu(p^2; p', m')} = \frac{\mu(p^1; p'', m'')}{\mu(p^2; p'', m'')}$$

This follows immediately from (28) and (11).

PROPOSITION H7: *Condition* (29) *implies that for any fixed income $\bar{m} > 0$, the indirect preference relation R^* satisfies* (22).

Setting $p^1 = p^0$, $p^2 = p$, $(p', m') = (p, m)$ and $(p'', m'') = (p^0, m^0)$, and making use of property M3 of μ, (29) yields

$$(30)\quad V_{p^0}(p, m) = \mu(p^0; p, m)$$

$$= m^0 \frac{m}{\mu(p; p^0, m^0)} = m^0 V(p, m)$$

Since V_{p0} is an indirect utility function, the result follows from Proposition H1.

This completes the proof of the sufficiency of homotheticity of preferences for

---

[12] Note that (29) is essentially equivalent to a criterion of Samuelson and Subramanian Swamy (equation (2.5), p. 570), who in place of μ worked with the function $e(p; x)$ related to μ by $\mu(p; p', m') = e(p; h(p', m'))$. They noted that the analogous functional equation was characteristic of homothetic preferences. The condition is necessary and sufficient for $\mu(p; p^0, m^0)$ to be a "true cost-of-living index," in the sense that it furnishes the price index by which one should divide (deflate) the consumer's income m in order to obtain the indirect utility function (23).

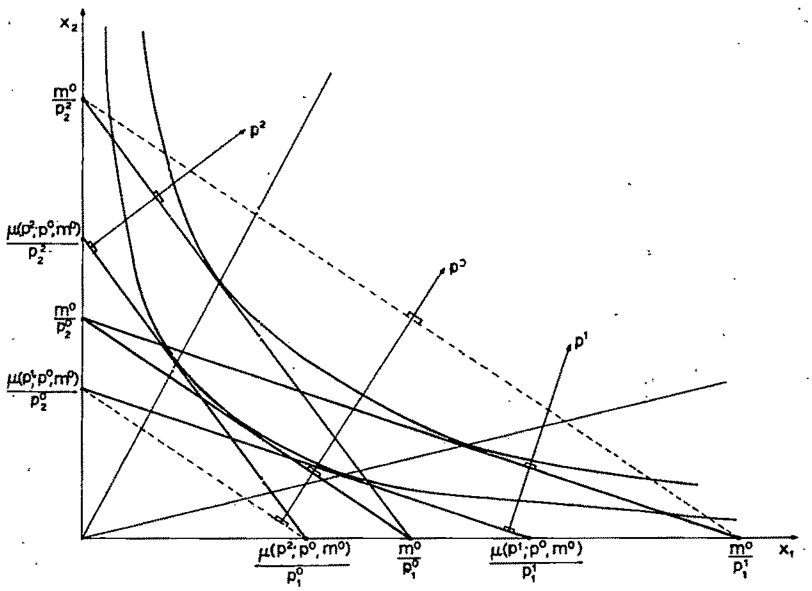


FIGURE 2

the compensating variation to be an acceptable measure of welfare change when income is constant and prices are free to vary arbitrarily. An illustration is given in Figure 2 of a case of homothetic preferences in which $(p^1, m^1)I^*(p^2, m^2)$, where $(p^1, m^1) = (p_1^1, p_2^0, m^0)$ and $(p^2, m^2) = (p_1^0, p_2^2, m^0)$. Since $(p^1, m^0)I^*(p^2, m^0)$ implies $\mu(p^1; p^0, m^0) = \mu(p^2; p^0, m^0)$ by (30), it follows that (see Figure 2)

$$\frac{m^0/p_2^0}{\mu(p^1;p^0,m^0)/p_2^0} = \frac{m^0/p_1^1}{\mu(p^1;p^0,m^0)/p_1^1}$$

$$= \frac{m^0/p_2^2}{\mu(p^2;p^0,m^0)/p_2^2} = \frac{m^0/p_1^0}{\mu(p^2;p^0,m^0)/p_1^0}$$

III. The Case of Fixed Price of Numéraire and Parallel Preferences

We now take up the case in which the price of a distinguished commodity (possibly a composite commodity) remains unchanged, or equivalently, income and all prices are reckoned throughout relative to the price of this distinguished commodity, which we take to be commodity 1. We show that for the generalized compensating variation to be an acceptable measure of welfare change in this case, it is necessary and sufficient that preferences be parallel with respect to commodity 1.

Let $\bar{p}_1 > 0$ be the fixed price of commodity 1, and denote $q = (p_2, p_3, \ldots, p_n)$ for convenience. Then the generalized compensating variation (17) correctly represents the consumer's indirect preference relation R^*, on a budget domain with $(p_1, q, m) = (\bar{p}_1, q, m)$, iff there is some base budget (p^0, m^0) such that

$$(31) \quad (\bar{p}_1, q^1, m^1)R^*(\bar{p}_1, q^2, m^2)$$

$$\text{iff } m^1 - \mu(\bar{p}_1, q^1; p^0, m^0)$$

$$\geqslant m^2 - \mu(\bar{p}_1, q^2; p^0, m^0)$$

As in the previous case, our demonstration will be divided into a series of seven propositions.

PROPOSITION P1: *Condition* (31) *is satisfied if and only if indirect preferences are* *representable by the function*

$$(32)\quad V(p,m)=\frac{m}{p_1}-\frac{\mu(p;p^0,m^0)}{p_1}$$

on the homogeneously extended budget domain.[13]

If $(p^1,m^1)=(p_1^1,q^1,m^1)$ and $(p^2,m^2)=(p_1^2,q^2,m^2)$ are any two budgets such that $(p^1,m^1)R^*(p,m^2)$, then by the homogeneity of degree zero of R^* we have $\lambda_1(p^1,m^1)$ $R^*\lambda_2(p^2,m^2)$ for $\lambda_1,\lambda_2>0$. Choosing $\lambda_1=\bar{p}_1/p_1^1$ and $\lambda_2=\bar{p}_1/p_1^2$ this yields

$$(\bar{p},\lambda_1 q^1,\lambda_1 m^1)R^*(\bar{p}_1,\lambda_2 q^2,\lambda_2 m^2)$$

hence we can apply the criterion (31). Using property M2 of μ this is easily seen to yield

$$\frac{m_1}{p_1^1}-\frac{\mu(p^1;p^0,m^0)}{p_1^1}\geqslant\frac{m_2}{p_1^2}-\frac{\mu(p^2;p^0,m^0)}{p_1^2}$$

Reversing the steps of the argument we see that the latter inequality implies (p^1,m^1) $R^*(p^2,m^2)$. Thus, $(p^1,m^1)R^*(p^2,m^2)$ iff

$$\frac{m^1}{p_1^1}-\frac{\mu(p^1;p^0,m^0)}{p_1^1}\geqslant\frac{m^2}{p_1^2}-\frac{\mu(p^2;p^0,m^0)}{p_1^2}$$

PROPOSITION P2: *If* R^* *is representable by* (32) *on an open budget domain, the demand function h satisfies*

$$(33)\quad \begin{aligned} h^1(p,m+\lambda p_1)&=h^1(p,m)+\lambda\\ h^j(p,m+\lambda p_1)&=h^j(p,m)\\ &\quad \text{for } j=2,3,\dots,n\end{aligned}$$

on this domain, and hence the preference relation R is parallel with respect to commodity 1.

This follows by applying the Antonelli equations (6) to (32) and using (13), to obtain

$$(34)\quad \begin{aligned} h^1(p,m)&=g^1(p;p^0,m^0)\\ &\quad +p_1^{-1}\left[m-\mu(p;p^0,m^0)\right]\\ h^j(p,m)&=g^j(p;p^0,m^0)\\ &\quad \text{for } j=2,3,\dots,n\end{aligned}$$

from which (33) immediately follows.

The constraint $h^1(p,m)\geqslant 0$ implies that the demand function can have the form (34) only for budgets (p,m) satisfying $m\geqslant\nu(p)$, where (recalling (13) and property M2 of μ)

$$(35)\quad \begin{aligned}\nu(p)&=\mu(p;p^0,m^0)-p_1 g^1(p;p^0,m^0)\\ &=\sum_{j=2}^{n}p_j g^j(p;p^0,m^0)\end{aligned}$$

The last $n-1$ equations of (34) further imply that the uncompensated and compensated demand functions are equal to one another, and depend only on prices, for all commodities but the first.[14] We may thus denote

$$(36)\quad \eta^j(p)=h^j(p,m)=g^j(p;p^0,m^0)\qquad \text{for } j=2,3,\dots,n,\qquad m\geqslant\nu(p)$$

The fulfillment of (34) for $m\geqslant\nu(p)$ can be shown to be necessary and sufficient for preferences to be parallel with respect to commodity 1 (see Samuelson, 1942, and our 1976 and 1979 papers).

Propositions P1 and P2 together imply that parallelism of preferences (with respect to commodity 1) is a necessary condition for the compensating variation to be an acceptable measure of welfare change when the price of commodity 1 is constant (or income and all other prices are reckoned relative to this price), but other prices and income are free to vary arbitrarily. The sufficiency follows from the remaining five propositions.

---

[13] As in fn. 11 above, if the original budget domain consists of budgets $(\bar{p}_1,\chi,\mu)$ such that the vector (χ,μ) varies in the set Θ, then the homogeneously extended domain is the set of budgets $(\lambda\bar{p}_1,\lambda\chi,\lambda\mu)$ with $\lambda>0$ and (χ,μ) in Θ, or equivalently, the set of budgets (p_1^1,q,m) such that $(\bar{p}_1/p_1)$ (q,m) is in Θ. See our 1979 paper.

[14] The fact that the property of "constancy" of the marginal utility of income $\partial V/\partial m=1/p_1$ enjoyed by (32) implies that the demand functions for the remaining commodities are independent of income was first pointed out by Samuelson (1942, p. 85). The fact that the uncompensated and compensated demand functions for these commodities are identical in this case was pointed out by Don Patinkin, p. 94.

PROPOSITION P3: *If the consumer's preference relation R is parallel with respect to commodity* 1, *then the indirect preference relation R^* satisfies*

$$(37)\qquad (p^1,m^1)R^*(p^2,m^2) \text{ iff }$$

$$(p^1,m^1+\lambda p_1^1)\ R^*(p^2,m^2+\lambda p_1^2)$$

$$\textit{for}\quad \lambda \geqslant \max\left\{\frac{\nu(p^1)-m^1}{p_1^1},\frac{\nu(p^2)-m^2}{p_1^2}\right\}$$

$$\textit{and}\quad m^t \geqslant \nu(p^t)\ \textit{for}\ t=1,2$$

Since (as noted above), parallelism implies (33), the result follows from parallelism, (33), and the definition (1) of R^*.

PROPOSITION P4: *If an indirect preference relation R^* satisfies* (37), *then for any fixed price vector p^t the Hurwicz-Uzawa function* (11) *has the property*

$$(38)\quad V_{p^t}(p,m+\lambda p_1)=V_{p^t}(p,m)+\lambda p_1^t$$

for p,m,λ, and $\mu(p^t;p,m)$ satisfying the inequalities in (37) *with $(p^t;\mu(p^t;p,m))$ and (p,m) replacing (p^1,m^1) and (p^2,m^2), respectively.*

By M4 we have $(p^t,\mu(p^t;p,m))I^*(p,m)$, whence it follows from Proposition P3 that $(p^t,\mu(p^t;p,m)+\lambda p_1^t)I^*(p,m+\lambda p_1)$. Applying M4 once again we obtain $\mu(p^t;p,m+\lambda p_1)=\mu(p^t;p,m)+\lambda p_1^t$.

PROPOSITION P5: *Property* (38) *implies that for any two base price vectors p^1 and p^2, the corresponding Hurwicz and Uzawa indirect utility representations $V_{p^t}(t=1,2)$ are related by*

$$(39)\quad V_{p^1}(p,m)=\frac{p_1^1}{p_1^2}V_{p^2}(p,m)+V_{p^1}(p^2,\nu(p^2))-\frac{p_1^1}{p_1^2}\nu(p^2)$$

where ν is given by (35); *that is, given any two such representations for R^*, one of them is an increasing linear transformation of the other.*

From (11) and property M4 of μ, we have by (38):

$$\mu(p^1;p,m)=\mu(p^1;p^2,\mu(p^2;p,m))$$

$$=\mu\left(p^1;p^2,\nu(p^2)+p_1^2\frac{\mu(p^2;p,m)-\nu(p^2)}{p_1^2}\right)$$

$$=\mu(p^1;p^2,\nu(p^2))+p_1^1\frac{\mu(p^2;p,m)-\nu(p^2)}{p_1^2}$$

By (11), this is equivalent to (39).

PROPOSITION P6: *Condition*(39) *implies that μ satisfies the functional equation*[15]

$$(40)\quad \frac{\mu(p^1;p',m')}{p_1^1}-\frac{\mu(p^2;p',m')}{p_1^2}=\frac{\mu(p^1;p'',m'')}{p_1^1}-\frac{\mu(p^2;p'',m'')}{p_2^2}$$

This follows immediately from (39) and (11).

PROPOSITION P7: *Condition* (40) *implies that for any fixed price of commodity* $1,\bar{p}_1>0$,

---

[15] Note that (40) provides conditions for $\mu(p;p^0,m^0)$ to be an ideal cost-of-living index in the sense of furnishing a cost figure $\mu(p;p^0,m^0)$ which should be *subtracted* from the consumer's income in order to obtain an expression which, after division by the price of the numéraire p_1, yields the indirect utility function (32).

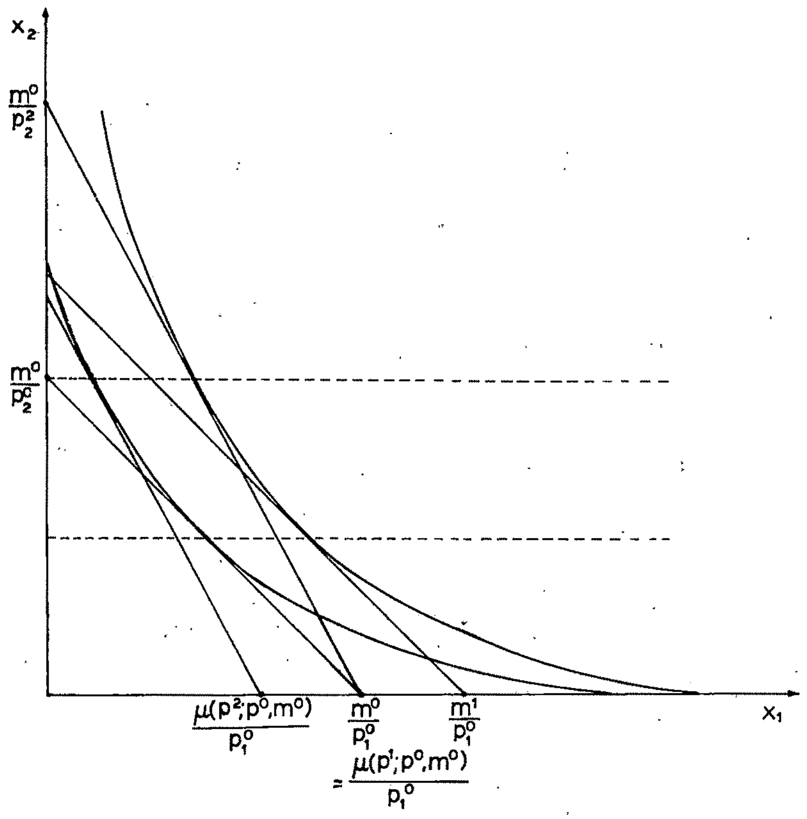


FIGURE 3

the indirect preference relation R^* *satisfies* (31) *over the relevant budget domain.*

Setting $p^1 = p^0, p^2 = p$, $(p', m') = (p, m)$, and $(p'', m'') = (p^0, m^0)$, and making use of property M3 of μ, (40) yields

$$(41)\quad V_{p^0}(p,m) = \mu(p^0;p,m) = m^0 + p_1^0\left[\frac{m}{p_1} - \frac{\mu(p;p^0,m^0)}{p_1}\right] = m^0 + p_1^0 V(p,m)$$

Since V_{p0} is an indirect utility function, the result then follows by Proposition P1.

This completes the proof of the sufficiency of parallelism of preferences for the compensating variation to be an acceptable measure of welfare change when one price is constant and income and the remaining prices are free to vary arbitrarily. An illustration is given in Figure 3 of a case of parallel preferences in which $(p^1, m^1)I^*(p^2, m^2)$, where $(p^1, m^1) = (p_1^0, p_2^0, m^1)$ and $(p^2, m^2) = (p_1^0, p_2^2, m^0)$. In this case, since $m^1 - \mu(p^1, p^0, m^0) = m^0 - \mu(p^2;p^0,m^0)$ by (41) it follows that (see Figure 3)

$$\frac{m^1}{p_1^0} - \frac{\mu(p^1;p^0,m^0)}{p_1^0} = \frac{m^0}{p_1^0} - \frac{\mu(p^2;p^0,m^0)}{p_1^0}.$$

where
$$\frac{m^0}{p_1^0} = \frac{\mu(p^1;p^0,m^0)}{p_1^0}$$

IV. The Derivation of Welfare Measures from Demand Functions

We conclude in this section with a discussion of how, in the light of the results of the preceding sections, one can most economically proceed from demand functions to welfare measures, and the extent to which and conditions under which the Dupuit-Marshall concept of consumer's surplus can serve a useful purpose or can be improved upon in fulfilling this objective. We first analyze the problem under the conditions assumed in Sections II and III (homothetic and parallel preferences), respectively, and then under more general conditions. We shall consider in particular the problem of deriving the Hurwicz and Uzawa function $V_{p0}(p,m)=\mu(p^0;p,m)$ either from the demand function $h(p,m)$, by integration of the appropriate Antonelli equations

$$(42)\quad v^j(p,m)=-\frac{\partial\mu(p^0;p,m)}{\partial m}h^j(p,m)\qquad \text{for } j=1,2,\dots,n$$

$$v^{n+1}(p,m)=\frac{\partial\mu(p^0;p,m)}{\partial m}$$

or from knowledge of some other indirect utility function $V(p,m)$ and use of property M4 of μ, that is , by solving the equation

$$(43)\quad V\left[p^0,\mu(p^0;p,m)\right]=V(p,m)$$

We shall also consider the problem of deriving the compensating (and equivalent) variations in the two cases we have considered in which the generalized compensating variation is an acceptable measure of welfare change.

A. *The Case of Homothetic Preferences*

Let preferences be homothetic, and consider first the case in which income is equal to a constant, $\bar{m}>0$. Defining the function f by (8), we see from the homotheticity of R and the definition (12) of g that

$$(44)\quad h^j(p,\bar{m})=\bar{m}h^j(p,1)=\frac{\bar{m}g^j(p;p^0,m^0)}{\mu(p;p^0,m^0)}$$

Using (13) we then verify that the function $F(\cdot\,;\bar{m})$ defined by

$$(45)\quad F(p;\bar{m})=\bar{m}\log\bar{m}-\bar{m}\log\mu(p;p^0,m^0)$$

is a potential function for the line integral of $\bar{f}=(\bar{f}^1,\bar{f}^2,\dots,\bar{f}^n)$ with respect to $p(\cdot)$, where $\bar{f}^j(p)=f^j(p,\bar{m})$. It follows, therefore, that the Marshallian consumer's surplus (9) corresponding to a path from $\omega(0)=(p^1,\bar{m})$ to $\omega(1)=(p^2,\bar{m})$ is equal to[16]

$$(46)\quad -\sum_{j=1}^{n}\int_0^1 h^j\left[p(t),\bar{m}\right]dp_j(t)$$

$$=\bar{m}\left[\log\mu(p^1;p^0,m^0)-\log\mu(p^2;p^0,m^0)\right]$$

On the other hand, the corresponding change in the generalized compensating variation (17), or equivalently, the Hicksian consumer's surplus (20) corresponding to a path from $\omega(0)=(p^1,\bar{m})$ to $\omega(1)=(p^2,\bar{m})$, is

$$(47)\quad -\sum_{j=1}^{n}\int_0^1 g^j\left[p(t);p^0,m^0\right]dp_j(t)$$

$$=\mu(p^1;p^0,m^0)-\mu(p^2;p^0,m^0)$$

Suppose now that $m^0=\bar{m}$, and consider the case in which p^1 and p^2 in (46) and (47) are replaced by p^0 and p^1, respectively. Denote the corresponding Marshallian consumer's surplus (9) by

$$(48)\quad S=-\int_0^1 h\left[p(t),m^0\right]dp(t)$$

and the corresponding equivalent and compensating variations (14) and (16) by

$$(49)\quad E=E(p^0;p^1,m^0)$$

$$\text{and}\quad C=C(p^1;p^0,m^0)$$

Then using M3 and (30) we see from (46),

---

[16] Despite appearances, the Marshallian consumer's surplus does not depend on the choice of base budget (p^0,m^0); such lack of dependence is precisely what is asserted by the functional equation (29).

(47), (14), and (16) that

$$(50) \qquad E/m^0 = e^{S/m^0} - 1$$

and

$$C/m^0 = 1 - e^{-S/m^0}$$

Thus, the equivalent and compensating variations can both be derived uniquely from the consumer's surplus, and all three correctly represent the indirect preferences of (p^1, m^0) over (p^0, m^0). Considering E and C in (50) as functions of S, by expanding in a Taylor's series around $S=0$ to one term we obtain the first-order approximations (as $S \to 0$)

$$(51) \quad E = S + o(S) \qquad C = S + o(S)$$

Moreover, by well-known properties of the exponential function we have $C < S < E$ for $0 < S/m^0 < 1$ (compare Hicks, 1956, p. 82). Note that while, in accordance with Robert Willig's results, both C and E can be approximated by S (if S is sufficiently small), they can be computed exactly. Moreover, S itself in this case is an acceptable measure of welfare change.

An alternative way of deriving C and E, in fact the generalized compensating and equivalent variations, is, however, available. We may verify that if the function w is defined by

$$(52) \quad w^j(p,m) = -m^{-1}h^j(p,m) = -h^j(p,1) = -\frac{g^j(p;p^0,m^0)}{\mu(p;p^0,m^0)} \qquad (j=1,2,\dots,n)$$

$$w^{n+1}(p,m) = m^{-1}$$

then the line integral of w with respect to ω is path independent and has the potential function

$$(53) \quad W(p,m) = \log m - \log \mu(p;p^0,m^0) = \log V(p,m)$$

where V is defined by (23). In view of (30), the generalized equivalent and compensating variations (15) and (17) are obtained from (53) by

$$(54) \qquad \Phi(p,m) = m^0\left[e^{W(p,m)} - 1\right]$$

and

$$\Psi(p,m) = m\left[1 - e^{-W(p,m)}\right]$$

Use of the integrand (52) is superior to use of (44) in that (52) furnishes an acceptable integral measure of welfare change without the need to assume that m is fixed. The line integral of (52), which is an increment of (53), may be considered a generalization of the Marshallian consumer's surplus, since it differs from that of (8) by the fact that the integrand has been multiplied by the appropriate integrating factor $\partial \log V/\partial m = 1/m$, where V is given by (23) (compare (7)).

Turning now to the system (42), there are two ways in which one could proceed to derive $\mu(p^0;p,m)$. One would be to integrate (52) to get (53) and thence (23) in order to obtain the function $\mu(\cdot,p^0,m^0)$, and then make use of the fact that $\partial\mu(p^0;p,m)/\partial m = m^0/\mu(p;p^0,m^0)$ by (30); one would finally integrate (42) to get $\mu(p^0;p,m)$. Alternatively, however, we may observe from (30) and (53) that $\mu(p^0;p,m) = m^0 \exp W(p,m)$ (see (54)), so the process of taking the line integral of (42) may be bypassed; one can instead take the line integral of (52) and apply the above transformation.

If, as is often the case in current econometric practice, one has a parametric form for an indirect utility function V, used to derive and estimate the demand function h via (6), the resulting parameter estimates will immediately enable one to estimate V. Assuming V to be homogeneous of degree one in m, one then obtains V_{p^0} from (43) by the formula $\mu(p^0;p,m) = V(p,m)/V(p^0,1)$.

B. *The Case of Parallel Preferences*

Let preferences be parallel with respect to commodity 1. Consider first the case in which the price of commodity 1 is equal to a constant, $\bar{p}_1>0$. In this case, defining the function f by (8), we see from (34) that

$$(55)\quad h^j(\bar{p}_1,q,m)=g^j(\bar{p}_1,q;p^0,m^0) \quad \text{for } j=2,3,\dots,n$$

where $q=(p_2,p_3,\dots,p_n)$, hence the line integral of $\tilde{f}=(\tilde{f}^2,\tilde{f}^3,\dots,\tilde{f}^{n+1})$ with respect to $[q(\cdot),m(\cdot)]$, where $\tilde{f}(q,m)=f^j(\bar{p}_1,q,m)$ for j=2,3,...,n and $\tilde{f}^{n+1}(q,m)=1$, has the potential function $F(\bar{p},\cdot)$ defined by

$$(56)\quad F(\bar{p}_1,q,m)=m-\mu(\bar{p}_1,q;p^0,m^0).$$

Accordingly, the Marshallian consumer's surplus corresponding to a path from $\omega(0)=(\bar{p}_1,q^1,m^1)$ to $\omega(1)=(\bar{p}_1,q^2,m^2)$ is

$$(57)\quad -\sum_{j=2}^{n}\int_0^1 h^j[\bar{p}_1,q(t),m(t)]\,dq_j(t) + m^2-m^1=m^2-\mu(\bar{p}_1,q^2;p^0,m^0) - m^1+\mu(\bar{p}_1,q^1;p^0,m^0)$$

which is precisely the same as the increment in the generalized compensating variation (17); and if $\bar{p}_1=p_1^0$, it is also (by virtue of (41)) precisely equal to the increment in the generalized equivalent variation (15). In this case, therefore, all three measures coincide and no approximations are needed.[17]

As in the homothetic case, however, one can improve upon the above procedure, as follows. Define the function v by

$$(58)\quad \begin{aligned} v^1(p,m)&=-(p_1)^{-1}h^1(p,m)\\ &=-(p_1)^{-2}\left[m-\sum_{i=1}^{2}p_i\eta^i(p)\right]\\ v^j(p,m)&=-(p_1)^{-1}h^j(p,m)\\ &=-(p_1)^{-1}\eta^j(p) \quad \text{for } j=2,3,\dots,n\\ v^{n+1}(p,m)&=(p_1)^{-1} \end{aligned}$$

for $m>\nu(p)=\Sigma_{i=1}^n p_i\eta^i(p)$, where ν and η^j are defined by (35) and (36), respectively. Then the line integral of v with respect to ω is independent of path and has the potential function (32). In view of (41), the generalized equivalent and compensating variations are obtained from (32) by

$$(59)\quad \Phi(p,m)=p_1^0V(p,m)$$

$$\text{and}\quad \Psi(p,m)=p_1V(p,m)$$

Use of the integrand (58) is superior to use of (8) in that it furnishes an acceptable integral measure of welfare change without the need to assume that p_1 is constant. The line integral of (58), which is an increment of (32), is a generalization of the Marshallian consumer's surplus since it differs from the line integral of (8) by the fact that the integrand has been multiplied by the appropriate integrating factor $\partial V/\partial m=1/p_1$.

Turning to (42), since $\partial\mu(p^0;p,m)/\partial m=p_1^0/p_1$ by (41), the function v of (42) is just a multiple p_1^0 of the function v of (58), and no problems present themselves at all; the function V_{p^0} can be estimated directly from (41), as a linear transformation of the potential function for (58).

If we are given an indirect utility function V with the property $V(p,m+\lambda p_1)=V(p,m)+\lambda$, one readily obtains V_{p^0} from (43) by $\mu(p^0;p,m)=p_1V(p,m)+\nu(p^0)-p_1V(p^0,\nu(p^0))$.

C. *The General Case*

Formulas (54) and (59) above show that in the two cases which we have examined in this paper in which the generalized compensating variation is an acceptable measure of welfare change,[18] the generalized compensating and equivalent variations can both be derived from the demand function by means of an appropriate modification of consumer's surplus, or even more directly, without the need for any approximations. In general, unless one's interest were limited to ordinal comparisons of budgets (p, m) with the base budget (p^0, m^0), or to computations of actual taxes or subsidies needed as compensation, there would be no point in trying to compute, let alone approximate, the generalized compensating variation;[19] but the problem of computing (or approximating) the generalized equivalent variation, or—what amounts to the same thing—the Hurwicz-Uzawa income-compensation function (11), remains as interesting as before.

---

[17] This is the case of Hicks' (1956) zero "income effect," or equivalently, E. J. Mishan's zero "welfare effect"; see John Currie, John Murphy, and Andrew Schmitz, p. 762. In Hicks' words: "In order that the Marshall measure of consumer's surplus should be a good measure, one thing alone is needful—that the income effect should be small" (1956, p. 177). As we have seen, the same is true of the Hicks measure.

In some cases, a simple solution via (43) is available. For example, if demand is generated by a special form of the "translog" indirect utility function given by (see Dale Jorgenson, Lawrence Lau, and Thomas Stoker)

$$(60)\quad \log V(p,m)=(\log p)'\left[\alpha+\tfrac{1}{2}B\log p\right] - (\log m)\iota'\left[\alpha-B\log p\right]$$

where $(\log p)'=(\log p_1, \log p_2,\ldots, \log p_n)$ and $\iota'=(1,1,\ldots,1)$, and α is a column vector and B a matrix, one obtains immediately

$$(61)\quad \log\mu(p^0;p,m)=\Big\{(\log p^0)'\left[\alpha+\tfrac{1}{2}B\log p^0\right] - (\log p)'\left[\alpha+\tfrac{1}{2}B\log p\right] + (\log m)\iota'\left[\alpha-B\log p\right]\Big\} \Big/ \iota'\left[\alpha-B\log p^0\right]$$

In contrast, suppose one chooses a parametric form for the indirect utility function such as the indirect addilog of Hendrick Houthakker:[20]

$$(62)\quad V(p,m)=\sum_{i=1}^{n}\alpha_i\left(\frac{m}{p_i}\right)^{\beta_i} \qquad (\alpha_i\geqslant 0,\ 0\leqslant\beta_i<1)$$

From the Antonelli equations (6) we obtain the demand functions

$$(63)\quad h^j(p,m)=\frac{\alpha_j\beta_j m^{\beta_j}p_j^{-\beta_j-1}}{\sum_{i=1}^{n}\alpha_i\beta_i m^{\beta_i-1}p_i^{-\beta_i}} \qquad (j=1,2,\ldots,n)$$

(These do not exhibit unitary income elasticity as long as the β_i are distinct.) Once estimates of the $2n$ parameters α_i, β_i have been obtained from the data, the function (62) can be estimated directly.[21] In this case

---

[18] It should be mentioned that these cases (homothetic prefemeces and constant income, and parallel preferences and constant price of the numéraire) are certainly not the only ones in which the generalized compensating variation is an acceptable measure of welfare change. Suppose we consider the textbook case in which one price (say that of commodity 1) varies while all other prices and income remain constant. For the generalized compensating variation to be an acceptable measure of welfare change in this case we require

$$(p_1^1,\bar{q},\bar{m})R^*(p_1^2,\bar{q},\bar{m})$$

$$\text{iff } \bar{m}-\mu(p_1^1,\bar{q};p^0,m^0)>\bar{m}-\mu(p_1^2,\bar{q};p^0,m^0)$$

Since μ is an increasing function of p_1 (by (13)), the criterion is equivalent to $p_1^1<p_1^2$. Since a fall in one price necessarily enlarges the budget set, this property of R° holds automatically and no restrictions on preferences are required. The common practice of conducting theoretical analysis in terms of this one case may help explain why greater attention has nto been paid in the literature ot the constraints on preferences that need to be imposed for the conventional measures of welfare change to be acceptable in our sense.

[19] If one is really interested only in ordinal comparison of a budget (p, m) with the base budget (p^0, m^0), it is difficult to see why one would seek a particular cardinal indicator for this purpose.

[20] This case has been discussed by Hause, p. 1161.

[21] One could of course also compute the marginal utility of income by which one should multiply (63) so as to furnish the appropriate line integral, but this would be unnecessary.

one does not generally obtain a closed-form relationship between V and V_{p0}.[22] However, as an indicator of welfare V would surely be more reliable than an approximation via Marshallian consumer's surplus of the generalized equivalent variation (15).

If there are strong reasons for preferring a money measure of indirect utility (and the history of the subject tells us that this is undoubtedly the case),[23] what is needed is a practical way to compute the function $\mu(p^0;\cdot,\cdot)$ given the demand functions h^j (which are estimated so as to satisfy the Slutsky integrability conditions) or given an indirect utility function generating those demand functions. The theoretical possibility of computing such a function is not in question: its existence was proved by Hurwicz and Uzawa. All that is needed is a practical algorithm.

---

[22] One might nevertheless be able to obtain a good numerical approximation by use of (43), and this would presumably be a more straightforward and accurate method than one based on approximations of the type obtained by Willig.

[23] See Samuelson (1974) and Abram Bergson for discussions of this point.

REFERENCES

R. G. D. Allen, "The Marginal Utility of Money and Its Application," *Economica*, May 1933, *13*, 186–209.

G. B. Antonelli, *Sulla teoria matematica della economia politica*, Pisa 1886; English translation "On the Mathematical Theory of Political Economy," in John S. Chipman et al., eds., *Preferences, Utility, and Demand*, New York 1971, 332–64.

Tom W. Apostol, *Mathematical Analysis*, Reading 1957.

A. Bergson, "A Note on Consumer's Surplus," *J. Econ. Lit.*, Mar. 1975, *13*, 38–44.

K. E. Boulding, "The Concept of Economic Surplus," *Amer. Econ. Rev.*, Dec. 1945, *35*, 851–69.

John S. Chipman, "Homothetic Preferences and Aggregation," *J. Econ. Theory*, May 1974, *8*, 26–38.

______ **et al.,** *Preferences, Utility, and Demand*, New York 1971.

______ **and James C. Moore,** "The Scope of Consumer's Surplus Arguments," in Anthony M. Tang et al., eds., *Evolution, Welfare and Time in Economics: Essays in Honor of Nicholas Georgescu-Roegen*, Lexington 1976, 69–123.

______ **and** ______, "Compensating Variation as a Measure of Welfare Change," unpublished manuscript, Univ. Minnesota, Mar. 1979.

J. M. Currie, J. A. Murphy, and A. Schmitz, "The Concept of Economic Surplus and Its Use in Economic Analysis," *Econ. J.*, Dec. 1971, *81*, 741–99.

J. Dupuit, "On the Measurement of the Utility of Public Works," in Kenneth J. Arrow and Tibor Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969, 255–83.

F. Y. Edgeworth, "Osservazioni sulla teoria matematica dell'economia politica con riguardo speciale ai principi di economia di Alfredo Marshall," *Giornale degli Economisti*, Mar. 1891, *2*, 233–45.

A. C. Harberger, "The Measurement of Waste," *Amer. Econ. Rev. Proc.*, May 1964, *54*, 58–76.

______, "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *J. Econ. Lit.*, Sept. 1971, *9*, 785–97.

J. C. Hause, "The Theory of Welfare Cost Measurement," *J. Polit. Econ.*, Dec. 1975, *83*, 1145–82.

J. R. Hicks, *Value and Capital*, Oxford 1939.

______, "Consumers' Surplus and Index-Numbers," *Rev. Econ. Stud.*, Summer 1942, *9*, 126–37.

______, *A Revision of Demand Theory*, Oxford 1956.

H. S. Houthakker, "Additive Preferences," *Econometrica*, Apr. 1960, *28*, 244–57.

L. Hurwicz and H. Uzawa, "On the Integrability of Demand Functions," in John S. Chipman et al., eds., *Preferences, Utility, and Demand*, New York 1971, 114–48.

D. W. Jorgenson, L. J. Lau, and T. M. Stoker, "Welfare Comparison under Exact Aggregation," *Amer. Econ. Rev. Proc.*, May 1980, *70*, 268–72.

L. W. McKenzie, "Demand Theory without a Utility Index," *Rev. Econ. Stud.*, June 1957, *24*, 185–89.

Edwin Mansfield, *Microeconomics*, New York 1970; 2d ed., 1975.

Alfred Marshall, *Principles of Economics*, 8th ed., London 1920.

E. J. Mishan, "Rent as a Measure of Welfare Change," *Amer. Econ. Rev.*, June 1959, *49*, 386–95.

H. Mohring, "Alternative Welfare Gain and Loss Measures," *Western Econ. J.*, Dec. 1971, *9*, 349–68.

D. Patinkin, "Demand Curves and Consumer's Surplus," in Carl F. Christ et al., eds., *Measurement in Economics, Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford 1963, 83–112.

René Roy, *De l'utilité. Contribution à la théorie des choix*, Paris 1942.

P. A. Samuelson, "Constancy of the Marginal Utility of Income," in Oscar Lange et al., eds., *Studies in Mathematical Economics and Econometrics, In Memory of Henry Schultz*, Chicago 1942, 75–91.

______, "Principles of Efficiency—Discussion," *Amer. Econ. Rev. Proc.*, May 1964, *54*, 93–96.

______, "Complementarity—An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory," *J. Econ. Lit.*, Dec. 1974, *12*, 1255–89.

______ **and S. Swamy,** "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *Amer. Econ. Rev.*, Sept. 1974, *64*, 566–93.

E. Silberberg, "Duality and the Many Consumer's Surpluses," *Amer. Econ. Rev.*, Dec. 1972, *62*, 942–52.

R. D. Willig, "Consumer's Surplus Without Apology," *Amer. Econ. Rev.*, Sept. 1976, *66*, 589–97.

Scale Economies, Product Differentiation, and the Pattern of Trade

By PAUL KRUGMAN*

For some time now there has been considerable skepticism about the ability of comparative cost theory to explain the actual pattern of international trade. Neither the extensive trade among the industrial countries, nor the prevalence in this trade of two-way exchanges of differentiated products, make much sense in terms of standard theory. As a result, many people have concluded that a new framework for analyzing trade is needed.[1] The main elements of such a framework—economies of scale, the possibility of product differentiation, and imperfect competition—have been discussed by such authors as Bela Balassa, Herbert Grubel (1967, 1970), and Irving Kravis, and have been "in the air" for many years. In this paper I present a simple formal analysis which incorporates these elements, and show how it can be used to shed some light on some issues which cannot be handled in more conventional models. These include, in particular, the causes of trade between economies with similar factor endowments, and the role of a large domestic market in encouraging exports.

The basic model of this paper is one in which there are economies of scale in production and firms can costlessly differentiate their products. In this model, which is derived from recent work by Avinash Dixit and Joseph Stiglitz, equilibrium takes the form of Chamberlinian monopolistic competition: each firm has some monopoly power, but entry drives monopoly profits to zero. When two imperfectly competitive economies of this kind are allowed to trade, increasing returns produce trade and gains from trade even if the economies have identical tastes, technology, and factor endowments. This basic model of trade is presented in Section I. It is closely related to a model I have developed elsewhere; in this paper a somewhat more restrictive formulation of demand is used to make the analysis in later sections easier.

The rest of the paper is concerned with two extensions of the basic model. In Section II, I examine the effect of transportation costs, and show that countries with larger domestic markets will, other things equal, have higher wage rates. Section III then deals with "home market" effects on trade patterns. It provides a formal justification for the commonly made argument that countries will tend to export those goods for which they have relatively large domestic markets.

This paper makes no pretense of generality. The models presented rely on extremely restrictive assumptions about cost and utility. Nonetheless, it is to be hoped that the paper provides some useful insights into those aspects of international trade which simply cannot be treated in our usual models.

I. The Basic Model

A. *Assumptions of the Model*

There are assumed to be a large number of potential goods, all of which enter symmetrically into demand. Specifically, we assume that all individuals in the economy have the same utility function,

$$U=\sum_i c_i^{\theta} \qquad 0<\theta<1 \tag{1}$$

where c_i is consumption of the ith good. The number of goods actually produced, n, will be assumed to be large, although smaller than the potential range of products.[2]

---

*Yale University and Massachusetts Institute of Technology.

[1] A paper which points out the difficulties in explaining the actual pattern of world trade in a comparative cost framework is the study of Gary Hufbauer and John Chilas.

There will be assumed to be only one factor of production, labor. All goods will be produced with the same cost function:

$$(2) \qquad l_i=\alpha+\beta x_i \qquad \alpha,\beta>0$$
$$i=1,\dots,n$$

where l_i is labor used in producing the ith good and x_i is output of that good. In other words, I assume a fixed cost and constant marginal cost. Average cost declines at all levels of output, although at a diminishing rate.

Output of each good must equal the sum of individual consumptions. If we can identify individuals with workers, output must equal consumption of a representative individual times the labor force:

$$(3) \qquad x_i=Lc_i \qquad i=1,\dots,n$$

We also assume full employment, so that the total labor force must just be exhausted by labor used in production:

$$(4) \qquad L=\sum_{i=1}^{n}(\alpha+\beta x_i)$$

Finally, we assume that firms maximize profits, but that there is free entry and exit of firms, so that in equilibrium profits will always be zero.

B. *Equilibrium in a Closed Economy*

We can now proceed to analyze equilibrium in a closed economy described by the assumptions just laid out. The analysis proceeds in three stages. First I analyze consumer behavior to derive demand functions. Then profit-maximizing behavior by firms is derived, treating the number of firms as given. Finally, the assumption of free entry is used to determine the equilibrium number of firms.

The reason that a Chamberlinian approach is useful here is that, in spite of imperfect competition, the equilibrium of the model is determinate in all essential respects because the special nature of demand rules out strategic interdependence among firms. Because firms can costlessly differentiate their products, and all products enter symmetrically into demand, two firms will never want to produce the same product; each good will be produced by only one firm. At the same time, if the number of goods produced is large, the effect of the price of any one good on the demand for any other will be negligible. The result is that each firm can ignore the effect of its actions on other firms' behavior, eliminating the indeterminacies of oligopoly.

Consider, then, an individual maximizing (1) subject to a budget constraint. The first-order conditions from that maximum problem have the form

$$(5) \qquad \theta c_i^{\theta-1}=\lambda p_i \qquad i=1,\dots,n$$

where p_i is the price of the ith good and λ is the shadow price on the budget constraint, that is, the marginal utility of income. Since all individuals are alike, (5) can be rearranged to show the demand curve for the ith good, which we have already argued is the demand curve facing the single firm producing that good:

$$(6) \qquad p_i=\theta\lambda^{-1}(x_i/L)^{\theta-1} \qquad i=1,\dots,n$$

Provided that there are a large number of goods being produced, the pricing decision of any one firm will have a negligible effect on the marginal utility of income. In that case, (6) implies that each firm faces a demand curve with an elasticity of $1/(1-\theta)$, and the profit-maximizing price is therefore

$$(7) \qquad p_i=\theta^{-1}\beta w \qquad i=1,\dots,n$$

where w is the wage rate, and prices and wages can be defined in terms of any (common!) unit. Note that since θ, β, and w are the same for all firms, prices are the same for all goods and we can adopt the shorthand $p=p_i$ for all i.

---

[2] To be fully rigorous, we would have to use the concept of a continuum of potential products.

The price p is independent of output given the special assumptions about cost and utility (which is the reason for making these particular assumptions). To determine profitability, however, we need to look at output. Profits of the firm producing good i are

$$(8)\qquad \pi_i=px_i-\{\alpha+\beta x_i\}w \qquad i=1,\ldots,n$$

If profits are positive, new firms will enter, causing the marginal utility of income to rise and profits to fall until profits are driven to zero. In equilibrium, then $\pi=0$, implying for the output of a representative firm:

$$(9)\qquad x_i=\alpha/(p/w-\beta)=\alpha\theta/\beta(1-\theta) \qquad i=1,\ldots,n$$

Thus output per firm is determined by the zero-profit condition. Again, since α, β, and θ are the same for all firms we can use the shorthand $x=x_i$ for all i.

Finally, we can determine the number of goods produced by using the condition of full employment. From (4) and (9), we have

$$(10)\qquad n=\frac{L}{\alpha+\beta x}=\frac{L(1-\theta)}{\alpha}$$

C. *Effects of Trade*

Now suppose that two countries of the kind just analyzed open trade with one another at zero transportation cost. To make the point most clearly, suppose that the countries have the same tastes and technologies; since we are in a one-factor world there cannot be any differences in factor endowments. What will happen?

In this model there are none of the conventional reasons for trade; but there will nevertheless be both trade and gains from trade. Trade will occur because, in the presence of increasing returns, each good (i.e., each differentiated product) will be produced in only one country—for the same reasons that each good is produced by only one firm. Gains from trade will occur because the world economy will produce a greater diversity of goods than would either country alone, offering each individual a wider range of choice.

We can easily characterize the world economy's equilibrium. The symmetry of the situation ensures that the two countries will have the same wage rate, and that the price of any good produced in either country will be the same. The number of goods produced in each country can be determined from the full-employment condition

$$(11)\qquad n=L(1-\theta)/\alpha; \qquad n^*=L^*(1-\theta)/\alpha$$

where L^* is the labor force of the second country and n^* the number of goods produced there.

Individuals will still maximize the utility function (1), but they will now distribute their expenditure over both the n goods produced in the home country and the n^* goods produced in the foreign country. Because of the extended range of choice, welfare will increase even though the "real wage" w/p (i.e., the wage rate in terms of a representative good) remains unchanged. Also, the symmetry of the problem allows us to determine trade flows. It is apparent that individuals in the home country will spend a fraction $n^*/(n+n^*)$ of their income on foreign goods, while foreigners spend $n/(n+n^*)$ of their income on home country products. Thus the value of home country imports measured in wage units is $Ln^*/(n+n^*)=LL^*/(L+L^*)$. This equals the value of foreign country imports, confirming that with equal wage rates in the two countries we will have balance-of-payments equilibrium.

Notice, however, that while the *volume* of trade is determinate, the *direction* of trade—which country produces which goods—is not. This indeterminacy seems to be a general characteristic of models in which trade is a consequence of economies of scale. One of the convenient features of the models considered in this paper is that nothing important hinges on who produces what within a group of differentiated products. There is an indeterminacy, but it doesn't matter. This result might not hold up in less special models.

Finally, I should note a peculiar feature of the effects of trade in this model. Both before and after trade, equation (9) holds; that is, there is no effect of trade on the scale of production, and the gains from trade come solely through increased product diversity. This is an unsatisfactory result. In another paper I have developed a slightly different model in which trade leads to an increase in scale of production as well as an increase in diversity.[3] That model is, however, more difficult to work with, so that it seems worth sacrificing some realism to gain tractability here.

II. Transport Costs

In this section I extend the model to allow for some transportation costs. This is not in itself an especially interesting extension although the main result—that the larger country will, other things equal, have the higher wage rate—is somewhat surprising. The main purpose of the extension is, however, to lay the groundwork for the analysis of home market effects in the next section. (These effects can obviously occur only if there are transportation costs.) I begin by describing the behavior of individual agents, then analyze the equilibrium.

A. *Individual Behavior*

Consider a world consisting of two countries of the type analyzed in Section I, able to trade but only at a cost. Transportation costs will be assumed to be of the "iceberg" type, that is, only a fraction g of any good shipped arrives, with $1-g$ lost in transit. This is a major simplifying assumption, as will be seen below.

An individual in the home country will have a choice over n products produced at home and n^* products produced abroad. The price of a domestic product will be the same as that received by the producer p. Foreign products, however, will cost more than the producer's price; if foreign firms charge p^*, home country consumers will have to pay the c.i.f. price $\hat{p}^*=p^*/g$. Similarly, foreign buyers of domestic products will pay $\hat{p}=p/g$.

Since the prices to consumers of goods of different countries will in general not be the same, consumption of each imported good will differ from consumption of each domestic good. Home country residents, for example, in maximizing utility will consume $(p/\hat{p}^*)^{1/(1-\theta)}$ units of a representative imported good for each unit of a representative domestic good they consume.

To determine world equilibrium, however, it is not enough to look at consumption; we must also take into account the quantities of goods used up in transit. If a domestic resident consumes one unit of a foreign good, his combined direct and indirect demand is for $1/g$ units. For determining total demand, then, we need to know the ratio of total demand by domestic residents for each foreign product to demand for each domestic product. Letting σ denote this ratio, and σ^* the corresponding ratio for the other country, we can show that

$$\sigma=(p/p^*)^{1/(1-\theta)}g^{\theta/(1-\theta)} \tag{12}$$

$$\sigma^*=(p/p^*)^{-1/(1-\theta)}g^{\theta/(1-\theta)}$$

The overall demand pattern of each individual can then be derived from the requirement that his spending just equal his wage; that is, in the home country we must have $(np+\sigma n^*p^*)d=w$, where d is the consumption of a representative domestic good; and similarly in the foreign country.

This behavior of individuals can now be used to analyze the behavior of firms. The important point to notice is that the elasticity of *export* demand facing any given firm is $1/(1-\theta)$, which is the same as the elasticity of *domestic* demand. Thus transportation costs have no effect on firms' pricing policy; and the analysis of Section I can be carried out as before, showing that transportation costs also have no effect on the number of firms or output per firm in either country.

---

[3] To get an increase in scale, we must assume that the demand facing each individual firm becomes more elastic as the number of firms increases, whereas in this model the elasticity of demand remains unchanged. Increasing elasticity of demand when the variety of products grows seems plausible, since the more finely differentiated are the products, the better substitutes they are likely to be for one another. Thus an increase in scale as well as diversity is probably the "normal" case. The constant elasticity case, however, is much easier to work with, which is my reason for using it in this paper.

Writing out these conditions again, we have

$$(13)\quad p=w\beta/\theta;\ p^*=w^*\beta/\theta$$

$$n=L(1-\theta)/\alpha;\ n^*=L^*(1-\theta)/\alpha$$

The only way in which introducing transportation costs modifies the results of Section I is in allowing the possibility that wages may not be equal in the two countries; the number and size of firms are not affected. This strong result depends on the assumed form of the transport costs, which shows at the same time how useful and how special the assumed form is.

B. *Determination of Equilibrium*

The model we have been working with has a very strong structure—so strong that transport costs have no effect on either the numbers of goods produced in the countries, n and n^*, or on the prices relative to wages, p/w and p^*/w^*. The only variable which can be affected is the relative wage rate $w/w^*=\omega$, which no longer need be equal to one.

We can determine ω by looking at any one of three equivalent market-clearing conditions: (i) equality of demand and supply for home country labor; (ii) equality of demand and supply for foreign country labor; (iii) balance-of-payments equilibrium. It will be easiest to work in terms of the balance of payments. If we combine (12) with the other equations of the model, it can be shown that the home country's balance of payments, measured in *wage units* of the *other* country, is

$$(14)\quad B=\frac{\sigma^*n\omega}{\sigma^*n+n^*}L^*-\frac{\sigma n^*}{n+\sigma n^*}\omega L$$

$$=\omega LL^*\left[\frac{\sigma^*}{\sigma^*L+L^*}-\frac{\sigma}{L+\sigma L^*}\right]$$

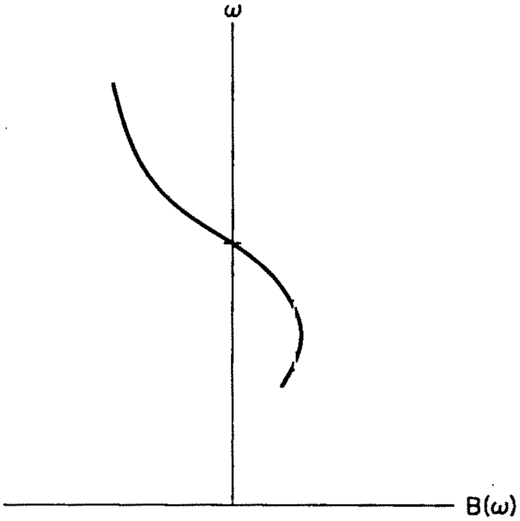


FIGURE 1

Since σ and σ^* are both functions of $p/p^*=\omega$, the condition $B=0$ can be used to determine the relative wage. The function $B(\omega)$ is illustrated in Figure 1. The relative wage $\bar{\omega}$ is that relative wage at which the expression in brackets in (4) is zero, and at which trade is therefore balanced. Since σ is an increasing function of ω and σ^* a decreasing function of ω, $B(\omega)$ will be negative (positive) if and only if ω is greater (less) than $\bar{\omega}$, which shows that $\bar{\omega}$ is the unique equilibrium relative wage.

We can use this result to establish a simple proposition: *that the larger country, other things equal, will have the higher wage.* To see this, suppose that we were to compute $B(\omega)$ for $\omega=1$. In that case we have $\sigma=\sigma^*<1$. The expression for the balance of payments reduces to

$$(14')\quad B=LL^*\left[\frac{1}{\sigma L+L^*}-\frac{1}{L+\sigma L^*}\right]$$

But (14′) will be positive if $L>L^*$, negative if $L<L^*$. This means that the equilibrium relative wage ω must be greater than one if $L>L^*$, less than one if $L<L^*$.

This is an interesting result. In a world characterized by economies of scale, one would expect workers to be better off in larger economies, because of the larger size of the local market. In this model, however, there is a secondary benefit in the form of better terms of trade with workers in the rest of the world. This does, on reflection, make intuitive sense. If production costs were the same in both countries, it would always be more profitable to produce near the larger market, thus minimizing transportation costs. To keep labor employed in both countries, this advantage must be offset by a wage differential.

III. "Home Market" Effects on the Pattern of Trade

In a world characterized both by increasing returns and by transportation costs, there will obviously be an incentive to concentrate production of a good near its largest market, even if there is some demand for the good elsewhere. The reason is simply that by concentrating production in one place, one can realize the scale economies, while by locating near the larger market, one minimizes transportation costs. This point—which is more often emphasized in location theory than in trade theory—is the basis for the common argument that countries will tend to export those kinds of products for which they have relatively large domestic demand. Notice that this argument is wholly dependent on increasing returns; in a world of diminishing returns strong domestic demand for a good will tend to make it an import rather than an export. But the point does not come through clearly in models where increasing returns take the form of external economies (see W. M. Corden). One of the main contributions of the approach developed in this paper is that by using this approach the home market can be given a simple formal justification.

I will begin by extending the basic closed economy model to one in which there are two industries (with many differentiated products within each industry). It will then be shown for a simple case that when two countries of this kind trade, each will be a net exporter in the industry for whose products it has the relatively larger demand. Finally, some extensions and generalizations will be discussed.

A. *A Two-Industry Economy*

As in Section I, we begin by analyzing a closed economy. Assume that there are two classes of products, *alpha* and *beta*, with many potential products within each class. A tilde will distinguish *beta* products from *alpha* products; for example, consumption of products in the first class will be represented as $c_1, \ldots, c_n$ while consumption of products in second are $\tilde{c}_1, \ldots, \tilde{c}_n$.

Demand for the two classes of products will be assumed to arise from the presence of two groups in the population.[4] There will be one group with L members, which derives utility only from consumption of *alpha* products; and another group with $\tilde{L}$ members, deriving utility only from *beta* products. The utility functions of representative members of the two classes may by written

$$(15)\quad U=\sum_i c_i^\theta;\ \tilde{U}=\sum_j \tilde{c}_j^\theta \qquad 0<\theta<1$$

For simplicity assume that not only the form of the utility function but the parameter θ is the same for both groups.

On the cost side, the two kinds of products will be assumed to have identical cost functions:

$$(16)\quad l_i=\alpha+\beta x_i \qquad i=1,\ldots,n$$

$$\tilde{l}_j=\alpha+\beta\tilde{x}_j \qquad j=1,\ldots,\tilde{n}$$

where, l_i, $\tilde{l}_j$ are labor used in production on typical goods in each class, and x_i, $\tilde{x}_j$ are total outputs of the goods.

The demand conditions now depend on the population shares. By analogy with (3), we have

---

[4] An alternative would be to have all people alike, with a taste for both kinds of goods. The results are similar. In fact, if each industry receives a fixed share of expenditure, they will be identical.

$$(17)\quad x_i = Lc_i \qquad i=1,\dots,n$$

$$\tilde{x}_j = \tilde{L}\tilde{c}_j \qquad j=1,\dots,\tilde{n}$$

The full-employment condition, however, applies to the economy as a whole:

$$(18)\qquad \sum_{i=1}^{n} l_i + \sum_{j=1}^{\tilde{n}} \tilde{l}_j = L + \tilde{L}$$

Finally, we continue to assume free entry, driving profits to zero. Now it is immediately apparent that the economy described by equations (15)–(18) is very similar to the economy described in equations (1)–(4). The price and output of a representative good—of either class—and the total number of products $n+\tilde{n}$ are determined just as if all goods belonged to a single industry. The only modification we must make to the results of Section I is that we must divide the total production into two industries. A simple way of doing this is to note that the sales of each industry must equal the income of the appropriate group in the population:

$$(19)\qquad npx = wL; \quad \tilde{n}\tilde{p}\tilde{x} = \tilde{w}\tilde{L}$$

But wages of the two groups must be equal, as must the prices and outputs of any products of either industry. So this reduces to the result $n/\tilde{n} = L/\tilde{L}$: the shares of the industries in the value of output equal the shares of the two demographic groups in the population.

This extended model clearly differs only trivially from the model developed in Section I when the economy is taken to be closed. When two such economies are allowed to trade, however, the extension allows some interesting results.

B. *Demand and the Trade Pattern: A Simple Case*

We can begin by considering a particular case of trade between a pair of two-industry countries in which the role of the domestic market appears particularly clearly. Suppose that there are two countries of the type just described, and that they can trade with transport costs of the type analyzed in Section II.

In the home country, some fraction f of the population will be consumers of *alpha* products. The crucial simplification I will make is to assume that the other country is a *mirror image* of the home country. The labor forces will be assumed to be equal, so that

$$(20)\qquad L+\tilde{L} = L^* + \tilde{L}^* = \bar{L}$$

But in the foreign country the population shares will be reversed, so that we have

$$(21)\qquad L = f\bar{L}; \quad L^* = (1-f)\bar{L}$$

If f is greater than one-half, then the home country has the larger domestic market for the *alpha* industry's products; and conversely. In this case there is a very simple home market proposition: *that the home country will be a net exporter of the first industry's products if* $f > 0.5$. This proposition turns out to be true.

The first step in showing this is to notice that this is a wholly symmetrical world, so that wage rates will be equal, as will the output and prices of all goods. (The case was constructed for that purpose.) It follows that the ratio of demand for each imported product to the demand for each domestic product is the same in both countries.

$$(22)\qquad \sigma = \sigma^* = g^{\theta/(1-\theta)} < 1$$

Next we want to determine the pattern of production. The expenditure on goods in an industry is the sum of domestic residents' and foreigners' expenditures on the goods, so we can write the expressions

$$(23)\quad npx = \frac{n}{n+\sigma n^*}wL + \frac{\sigma n}{\sigma n + n^*}wL^*$$

$$n^*px = \frac{\sigma n^*}{n+\sigma n^*}wL + \frac{n^*}{\sigma n + n^*}wL^*$$

where the price p of each product and the

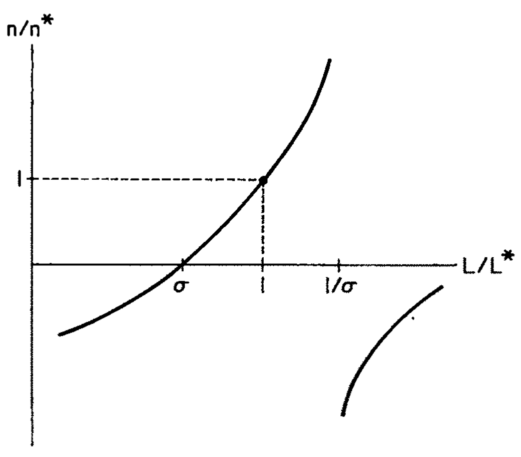


FIGURE 2

output x are the same in the two countries. We can use (23) to determine the relative number of products produced in each country, n/n^*.

To see this, suppose *provisionally* that some products in the *alpha* industry are produced in both countries; i.e., $n>0$, $n^*>0$. We can then divide the equations (23) through by n and n^*, respectively, and rearrange to get

$$L/L^* = (n+\sigma n^*)/(\sigma n + n^*) \tag{24}$$

which can be rearranged to give

$$n/n^* = \frac{L/L^* - \sigma}{1 - \sigma L/L^*} \tag{25}$$

Figure 2 shows the relationship (25). If $L/L^*=1$, so does n/n^*; that is, if the demand patterns of the two countries are the same, their production patterns will also be the same, as we would expect. And as the relative size of either country's home market rises for *alpha* goods, so does its domestic production, as long as L/L^* lies in the range $\sigma < L/L^* > 1/\sigma$.

Outside that range, (25) appears to give absurd results. Recall, however, that the derivation of (24) was made on the provisional assumption that n and n^* were both nonzero. Clearly, if L/L^* lies outside the range from σ to $1/\sigma$, this assumption is not valid. What the figure suggests is that if L/L^* is less than σ, $n=0$; the home country specialized entirely in *beta* products, producing no *alpha* products (while the foreign country produces only *alpha* products). Conversely, if L/L^* is greater than $1/\sigma$, $n^*=0$, and we have the opposite pattern of specialization.

We can easily demonstrate that this solution is in fact an equilibrium. Suppose that the home country produced no *alpha* products, and that a firm attempted to start production of a single product. This firm's profit-maximizing f.o.b. price would be the same as that of the foreign firm's. But its sales would be less, in the ratio

$$\frac{\sigma^{-1}L + \sigma L^*}{L + L^*} < 1$$

Thus such a firm could not compete.

This gives us our first result on the effect of the home market. It says that if the two countries have sufficiently dissimilar tastes each will specialize in the industry for which it has the larger home market. Obviously, also, each will be a net exporter of the class of goods in which it specializes. Thus the idea that the pattern of exports is determined by the home market is quite nicely confirmed.

We also get some illuminating results on the conditions under which specialization will be incomplete. Incomplete specialization and two-way trade within the two classes of products will occur if the relative size of the domestic markets for *alpha* goods lies in the range from σ to $1/\sigma$, where $\sigma = g^{\theta/(1-\theta)}$. But g measures transportation costs, while $\theta/(1-\theta)$ is, in equilibrium, the ratio of variable to fixed costs;[5] that is, it is an index of the importance of scale economies. So we have shown that the possibility of incomplete specialization is greater, the greater are transport costs and the less important are economies of scale.

A final result we can take from this special case concerns the pattern of trade when specialization is incomplete. In this case each country will both import and export products in *both* classes (though not the same products). But it remains true that, if one country has the larger home market for *alpha* producers, it will be a *net* exporter in the *alpha* class and a net importer in the other. To see this, note that we can write the home country's trade balance in *alpha* products as

---

[5] One can see this by rearranging equation (9) to get $\beta x/\alpha = \theta/(1-\theta)$.

$$(26)\quad B_\alpha = \frac{\sigma n}{\sigma n + n^*} wL^* - \frac{\sigma n^*}{n + \sigma n^*} wL$$
$$= wL^*\left[\frac{\sigma n}{\sigma n + n^*} - \frac{\sigma n^*}{n + \sigma n^*}\frac{L}{L^*}\right]$$
$$= \frac{\sigma w L^*}{\sigma n + n^*}\left[n - n^*\right]$$

where we used (24) to eliminate the relative labor supplies. This says that the sign of the trade balance depends on whether the number of *alpha* products produced in the home country is more or less than the number produced abroad. But we have already seen that n/n^* is an increasing function of L/L^* in the relevant range. So the country with the larger home market for the *alpha*-type products will be a net exporter of those goods, even if specialization is not complete.

C. *Generalizations and Extensions*

The analysis we have just gone through shows that there is some justification for the idea that countries export what they have home markets for. The results were arrived at, however, only for a special case designed to make matters as simple as possible. Our next question must be the extent to which these results generalize.

One way in which generalization might be pursued is by abandoning the "mirror image" assumption: we can let the countries have arbitrary populations and demand patterns, while retaining all the other assumptions of the model. It can be shown that in that case, although the derivations become more complicated, the basic home market result is unchanged. Each country will be a net exporter in the industry for whose goods it has a relatively larger demand. The difference is that wages will in general not be equal; in particular, smaller countries with absolutely smaller markets for both kinds of goods will have to compensate for this disadvantage with lower wages.

Another, perhaps more interesting, generalization would be to abandon the assumed symmetry between the industries. Again, we would like to be able to make sense of some arguments made by practical men. For example, is it true that large countries will have an advantage in the production and export of goods whose production is characterized by sizeable economies of scale? This is an explanation which is sometimes given for the United States' position as an exporter of aircraft.

A general analysis of the effects of asymmetry between industries would run to too great a length. We can learn something, however, by considering another special case. Suppose that the *alpha* production is the same as in our last analysis, but that the production of *beta* goods is characterized by *constant* returns to scale and perfect competition. For simplicity, also assume that *beta* goods can be transported costlessly.

It is immediately apparent that in this case the possibility of trade in *beta* products will ensure that wage rates are equal. But this in turn means that we can apply the analysis of Part B, above, to the *alpha* industry. Whichever country has the larger market for the products of that industry will be a net exporter of *alpha* products and a net importer of *beta* products. In particular: if two countries have the same composition of demand, the larger country will be a net exporter of the products whose production involves economies of scale.

The analysis in this section has obviously been suggestive rather than conclusive. It relies heavily on very special assumptions and on the analysis of special cases. Nonetheless, the analysis does seem to confirm the idea that, in the presence of increasing returns, countries will tend to export the goods for which they have large domestic markets. And the implications for the pattern of trade are similar to those suggested by Steffan Linder, Grubel (1970), and others.

REFERENCES

Bela Balassa, *Trade Liberalization Among Industrial Countries*, New York 1967.

W. M. Corden, "A Note on Economies of Scale, the Size of the Domestic Market and the Pattern of Trade," in I. A. McDougall and R. H. Snape, eds., *Studies in International Economics*, Amsterdam 1970.

A. Dixit and J. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *Amer. Econ. Rev.*, June 1977, *67*, 297–308.

H. Grubel, "Intra-Industry Specialization and the Pattern of Trade," *Can. J. Econ.*, Aug. 1967, *33*, 374–388.

______, "The Theory of Intra-Industry Trade," in I. A. McDougall and R. H. Snape, eds., *Studies in International Economics*, Amsterdam 1970.

G. Hufbauer and J. Chilas, "Specialization by Industrial Countries: Extent and Consequences," in Herbert Giersch, ed., *The International Division of Labor*, Tübingen 1974.

I. Kravis, "The Current Case for Import Limitations," in *United States Economic Policy in an Interdependent World, Commission on International Trade and Investment Policy*, Washington 1971.

P. Krugman, "Increasing Returns, Monopolistic Competition, and International Trade," *J. Int. Econ.*, Nov. 1979, *9*, 469–80.

Steffan Linder, *An Essay on Trade and Transformation*, New York 1961.

Exchange Rates and the Current Account

By RUDIGER DORNBUSCH AND STANLEY FISCHER*

This paper develops a model of exchange rate determination that integrates the roles of relative prices, expectations, and the assets markets, and emphasizes the relationship between the behavior of the exchange rate and the current account. In its orientation, it draws on and extends recent work, particularly by Pentti Kouri (1975; 1976), most closely related to this paper, and also by Jacob Frenkel and Carlos Rodriguez, Robert Flood, Charles Wilson, and Dornbusch (1975; 1976).

Early theories of the exchange rate singled out purchasing power parity or the current account as the chief determinants of the exchange rate. Work of a macroeconomic orientation, in particular the influential papers by Robert Mundell and J. Marcus Fleming, introduced capital mobility as an important aspect of exchange rate determination and presented a first formulation of the assets market view. In this theory, the exchange rate achieves a level such that the goods market clears and money demand equals money supply at the prevailing level of income and world interest rates.

Extensions of this approach have emphasized the possibility of international nominal interest differentials arising from exchange rate expectations. Coupled with assumptions about the formation of exchange rate expectations, in particular rational expectations, these formulations have introduced the idea of exchange rate dynamics and of overshooting arising from differential adjustment speeds in goods and assets markets.

In a parallel development, exchange rate theory has been extended to reintroduce the current account as an important determinant of the exchange rate. In this perspective, assets markets determine the exchange rate at a point in time, but the current account through its effect on net asset positions, and therefore on assets markets, determines the path of the exchange rate over time. The introduction of rational expectations merges the assets and current account theories of exchange rates because it can lead to a fully anticipated equilibrium path in which asset prices adjust in part to reflect future current account developments.

The present paper extends this literature in two respects. First we depart from the one-commodity model that has been used extensively. Instead we assume that the country to be analyzed produces a differentiated product, whose world relative price is endogenous. This is an important departure from earlier models because the extension implies that the behavior of the exchange rate is not exclusively determined by the behavior of equilibrium real money demand via a strict purchasing power parity link.

Our second extension, drawing on work by Fischer, Olivier Blanchard, and Wilson, is to consider the current effects of anticipated future disturbances. This question is important because it suggests patterns of exchange rate and current account behavior that do not flow from expectations-augmented current account models.

Section I develops the basic model for the case of static expectations. The extension to perfect foresight is developed in Section II. Anticipated future disturbances are analyzed in Section III. Concluding comments are contained in Section IV.

I. The Model with Static Expectations

In this section we analyze a small open economy that trades in goods and securities with the rest of the world. The home country is at full employment, prices are flexible, and output is given. The world demand for the home country's product depends on the relative price or the terms of trade. In the import market, the home country is small in the sense that the world price of importables is given in terms of foreign exchange.

---

*Massachusetts Institute of Technology. We are indebted to George Borts and an anonymous referee for their suggestions. Financial support was provided by grants from the National Science Foundation.

The asset menu of our model is severely restricted so as to concentrate on essentials. Domestic residents can hold domestic money or foreign interest-earning bonds.[1] The foreign bonds are "real" bonds promising to pay a unit of foreign output indefinitely.[2] The world interest rate is taken as given.

Short-run equilibrium in this economy corresponds to an exchange rate and a domestic price level—a terms of trade and a real value of cash balances, equivalently—such that the money market clears and the market for domestic goods is in equilibrium. The equilibrium is contingent on a given stock of external assets, and in turn it determines equilibrium saving and hence the current account imbalance and asset accumulation until a long-run equilibrium with a balanced current account is attained.

A. *Short-Run Equilibrium and Dynamics*

We start by setting out the basic model. Equilibrium in the money market obtains when money demand, a function of income and the alternative cost of holding money is equal to the supply of money:

$$(1) \qquad M=k(r^*+x)[Py+eP^*a] \qquad k'<0$$

where M is the nominal money supply, r^* and x are the given foreign interest rate and the anticipated rate of depreciation of the exchange rate, P is the price of domestic output, and y is domestic physical output. The term eP^*a denotes the income from foreign bonds measured in terms of domestic currency, which is of course part of domestic income. Here e is the exchange rate, P^* the given foreign price level, and a the number of income streams each yielding one unit of foreign output indefinitely.

Dividing equation (1) by the domestic price of domestic output we have one of our equilibrium conditions:

$$(1') \quad m=k[y+\lambda a];\, m\equiv M/P;\, \lambda\equiv eP^*/P$$

where m and λ denote the real value of money and the terms of trade, respectively. For the present we treat expected depreciation as identically equal to zero so that in conjunction with the given foreign interest rate the opportunity cost of holding money is constant and so accordingly is k.

In the goods market we assume that domestic demand for our output, D, is a function of the terms of trade λ, and real wealth w. Foreign demand for our goods, X, is just a function of the terms of trade:

$$(2) \quad y=D(\lambda,w)+X(\lambda);\, D_\lambda, X_\lambda>0;\, D_w>0$$

where real wealth is the sum of the real value of cash balances plus the real value of external assets, $\lambda a/r^*$:

$$(3) \qquad w\equiv m+\lambda a/r^*$$

An increase in the relative price of foreign goods (a rise in λ) is assumed to shift demand toward domestic goods. An increase in wealth raises aggregate spending, part of which falls on domestic output thus raising demand.

When the goods market clears, the current account equals the excess of income over spending. In our economy, without a government sector, taxes, or investment, the excess of income over spending is equal to saving. Saving is assumed to be a decreasing function of nonhuman wealth and an increasing function of labor income where the latter is by assumption constant and therefore suppressed as an argument:[3]

$$(4) \qquad S=S(w);\, S_w<0$$

[1] Without loss of generality, we exclude domestic holdings of foreign monies since they are dominated as securities by foreign interest-earning assets.

[2] It is inevitable that a special case has to be chosen —foreign bonds could be nominal or real, in domestic currency or foreign currency, in terms of domestic or foreign output. The choice does have implications for the results we derive.

[3] This saving behavior reflects the standard life cycle theory of saving and consumption. See Franco Modigliani. The reader will note an asymmetry between our money demand which is, along transactions

Finally we note that the current account surplus is equal to the rate at which the home country acquires claims on the rest of the world. Accordingly we can write

$$\dot{a}/r^* = S(w)/\lambda \tag{5}$$

This completes the formal model and we now proceed to show the determination of short-run equilibrium.

We can solve (1′) for the equilibrium stock of real balances as an increasing function of the value of external assets:

$$m = m(\lambda a) \qquad m' > 0 \tag{1''}$$

An increase in the value of external assets raises real income, real money demand, and therefore, via a decline in the price level, the equilibrium real money stock.

Having determined equilibrium real balances, we can turn to the goods market where we determine the equilibrium terms of trade as a function of the given level of external assets, a. With real balances determined in (1″) real wealth is a function of the terms of trade and external assets. An increase in external assets a raises real wealth directly and via the induced change in real balances. There is accordingly an increase in demand for domestic output. To restore equilibrium the terms of trade must improve —λ, the relative price of imports, must decline. The decline in the terms of trade falls proportionately short of the increase in assets. Otherwise real wealth would be unchanged or rise so that the relative price decline via the substitution effect would leave an excess demand.[4]

$$\lambda = \lambda(a);\ \lambda' < 0;\ \phi \equiv -(a/\lambda) d\lambda/da < 1 \tag{6}$$

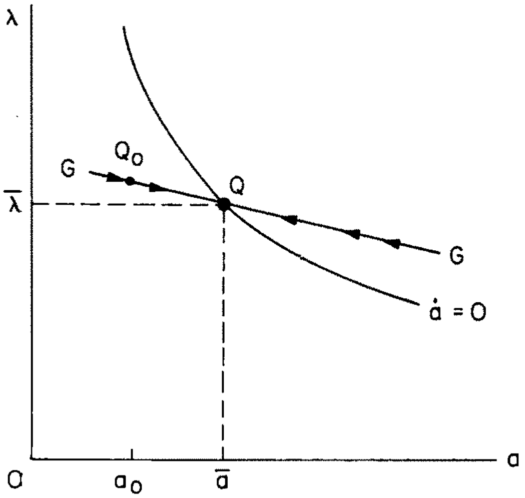


FIGURE 1

In Figure 1 we show as the GG schedule the relation between the equilibrium terms of trade and external assets. Moving down and along the schedule real wealth rises, raising demand for domestic output and therefore requiring an offsetting increase in the relative price of domestic goods.

We also show in Figure 1 the schedule $\dot{a}=0$ along which the current account is in balance. The schedule is a rectangular hyperbola since wealth is constant only for a given value of external assets. The schedule must cross the GG schedule from above. To the left of the $\dot{a}=0$ schedule, real wealth is too low, saving is positive, and therefore assets are being accumulated. Conversely, to the right of the $\dot{a}=0$ schedule, there is dissaving and asset decumulation through a current account deficit.

At any point in time there exists an amount of external assets, say a_o. Associated with that asset position there is an equilibrium stock of real balances and a level of real wealth that determine spending, the equilibrium terms of trade $\lambda(a_o)$, and the current account. At point Q_o in Figure 1, with assets a_o, there is a current account surplus and asset accumulation. The accumulation will persist until we reach the steady-state level of assets $\bar{a}$ where the current account achieves balance.

Note particularly that, as the economy moves along GG, the current account is in

---

demand lines, a function of income, and our saving function where wealth is a determinant. It is readily verified that introducing wealth rather than income as a determinant of money demand will not change the analysis. Income as the sole determinant of saving, on the contrary, would imply instability.

[4] From equations (1′), (2), and (3), we can calculate the expression for the proportional change in the terms of trade induced by a change in assets:

$$(a/\lambda)\, d\lambda/da = -\frac{(a/r^* + ka)\, D_w}{D_w(a/r^* + ka) + D_\lambda + X_\lambda}$$

Accordingly $\phi \equiv -(a/\lambda)\, d\lambda/da$ is less than unity in absolute value.

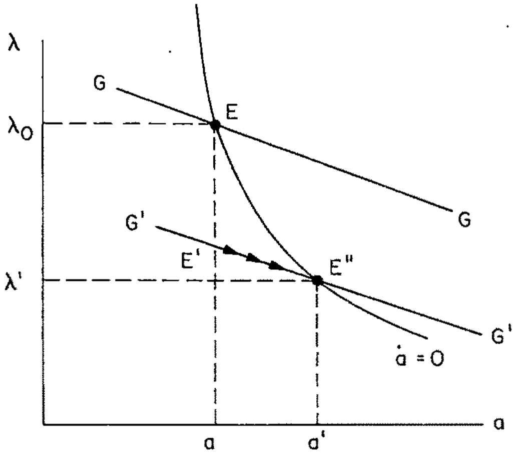


FIGURE 2

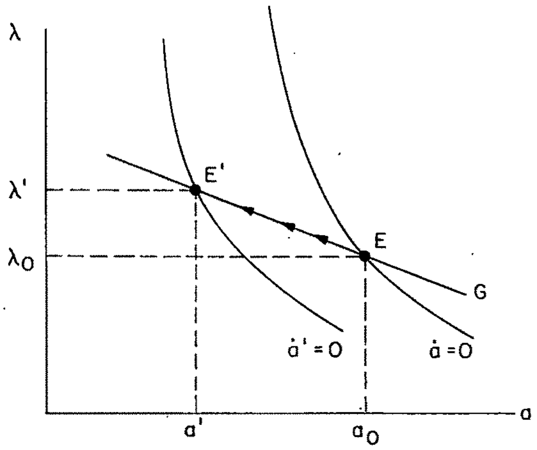


FIGURE 3

surplus (we are acquiring foreign assets). The terms of trade improve and with P^* fixed, a current account surplus accompanies an appreciating exchange rate.

To demonstrate the stability of the model formally, we observe the stability condition:

$$(7) \qquad \frac{d\dot{a}}{da} = S_w\left(1-\phi+\frac{rm'}{\lambda}\right)+\frac{\dot{a}}{a}\phi$$

As shown above, ϕ is less than one along the GG schedule so that stability is ensured in the neighborhood of the steady state.

B. *Comparative Statics*

The long-run equilibrium shown in Figure 1 can be disturbed by changes in exports or the composition of domestic spending, on one side, or by changes in saving with corresponding increases in the level of spending on either domestic goods or imports.

Consider first an increase in world demand for our goods, a shift in the $X(\lambda)$ function in equation (2). At the initial equilibrium there is now an excess demand for domestic output, and accordingly, we require an improvement in the terms of trade, or a decline in wealth through reduced external assets, to maintain market equilibrium. Thus in Figure 2 the GG schedule shifts to $G'G'$. The new long-run equilibrium is at E'' with an improvement in the terms of trade and an increase in external assets.

The adjustment process to the new equilibrium involves an immediate improvement in the terms of trade to point E', and at that point a current account surplus. The current account surplus leads to external asset accumulation and further improvement in the terms of trade until real wealth returns to its initial level at point E''. At E'' the trade deficit, measured in terms of foreign goods, increases by the increase in a. In terms of domestic goods the deficit is equal to λa and remains unchanged (between E and E'').

As an alternative disturbance we consider now an increase in imports arising from a reduction in saving. Assume that in terms of equation (4) we have a reduction in planned saving with a corresponding increase in import spending. In Figure 3 we show this as a leftward shift of the zero saving schedule. We now require lower external assets to maintain zero saving. The new long-run equilibrium at point E' is one with a decline in external assets and a worsening of the terms of trade. The new equilibrium is reached through the current account deficit at point E that leads to external dissaving. The demand for domestic goods declines with the reduction in external assets, and domestic competitiveness has to be enhanced through a decline in the relative price of domestic goods so as to maintain market equilibrium.

C. *Exchange Rate Behavior in the Adjustment Process*

We focus now on exchange rate movements in the adjustment process, and particularly on the relationship between the exchange rate and the current account. It is apparent that the exchange rate is determined along with the general equilibrium in the goods and assets markets. We can study its behavior by starting from the definition of the terms of trade, $\lambda \equiv eP^*/P$. Accordingly we can write the exchange rate as

$$(8)\quad e \equiv \lambda P/P^* \equiv [\lambda/(M/P)][M/P^*]$$

Equilibrium λ is determined from equation (2) (and shown by the *GG* schedule) and, given y, is a function only of a. Equilibrium M/P is from (1′) a function of a and λ, but since the latter is a function of a, M/P too may be taken to be a function only of a. Accordingly we can write the equilibrium exchange rate as a (diminishing)[5] function of a:

$$(9)\quad e = \frac{\lambda(a)}{M/P(\lambda, a)}(M/P^*) = J(a)(M/P^*) \qquad J_a < 0$$

Expanding (9) around the long-run equilibrium, we can write the linear approximation:

$$(10)\quad e = \bar{e} + \gamma(a - \bar{a});\ \gamma \equiv J_a(M/P^*) < 0$$

where the bar denotes long-run equilibrium values. Equation (10) shows that the equilibrium exchange rate is equal to the long-run equilibrium rate $\bar{e}$, adjusted for the deviation of external assets from their steady-state level.

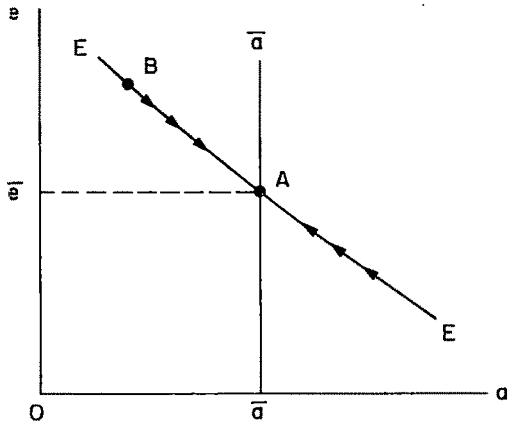


FIGURE 4

The equilibrium exchange rate path is shown in Figure 4 as the *EE* schedule. The figure shows that for levels of external assets below the steady-state level $\bar{a}$, such as at point *B*, the exchange rate is above its steady-state level $\bar{e}$. Assets are accumulated as a moves up to $\bar{a}$ (from point *B* to point *A*), at the same time as e appreciates to $\bar{e}$. Since asset accumulation takes place through a current account surplus, we have a current account surplus accompanying an appreciating exchange rate in the adjustment process. This is in accordance with the popular view of the relation between the exchange rate and the current account.[6]

II. Exchange Rate Expectations and the Adjustment Process

In this section, we take account of expectations of inflation in their effect on velocity and thus on the terms of trade, the exchange rate, and the rate of accumulation. Once again, we are in a full-employment context. We start with the recognition that the alternative assets are domestic money and foreign assets. By assumption, goods are not storable. Thus the relevant alternative cost to money holders is the foreign interest rate

---

[5] From (1″), (6), and (8), we have

$$de/da = -(e/a)\left[\phi + (1-\phi)\frac{a\lambda}{Y + a\lambda}\right] < 0$$

[6] It may be useful for the reader to go back to Figures 2 and 3 and confirm in those cases the relationship between the exchange rate and the current account specified in this paragraph.

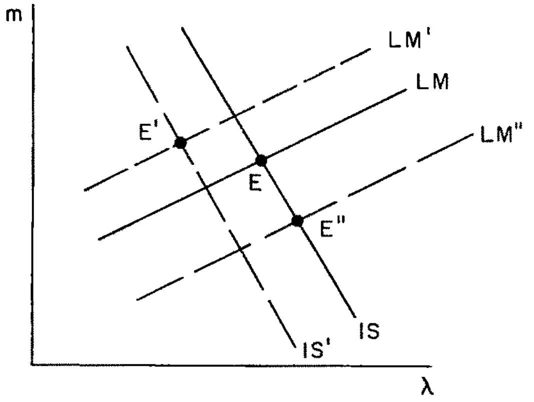


FIGURE 5

plus the anticipated rate of depreciation x. Given the foreign interest rate, velocity thus depends on the expected rate of depreciation. We are interested now in describing the adjustment process once these expectations, allowed to be rational, are taken into account.

We shall eventually study the dynamic adjustment process in some detail. But we begin with the short-run (instantaneous) equilibrium of the economy, implied by equations (1″) and (2), the money and goods market equilibrium conditions, respectively. Note that now $k=k(x)$[7] and that an increase in the anticipated rate of depreciation of our currency raises velocity or reduces k.

The instantaneous equilibrium determines the terms of trade λ and the stock of real balances m as functions of the stock of foreign assets a, and the expected rate of depreciation x. Figure 5 shows the money market equation (1′) and goods market equation (2) as the *LM* and *IS* curves, respectively. We repeat (1′) and (2) for convenience:

$$m=k(x)[y+\lambda a];\ m\equiv M/P \tag{1′}$$

$$y=D(\lambda, m+\lambda a/r^*)+X(\lambda) \tag{2}$$

The *LM* curve is upward sloping since an increase in λ increases income measured in domestic goods, and increases the demand for money. The *IS* curve slopes downwards because an increase in λ increases the demand for our output, and has therefore to be offset by a fall in m to maintain goods market equilibrium.

The equilibrium at E depends, at any moment, on the values of a and x in a manner we shall now examine. An increase in assets a increases the demand for money and moves the *LM* curve up to *LM′*; it also increases the demand for goods and therefore moves the *IS* curve down to *IS′* (a lower m is needed to maintain goods market equilibrium). The new equilibrium is at E'. The terms of trade unambiguously improve, while it can be shown that the stock of real balances rises, that is, that the horizontal shift in the *IS* curve is less than that of the *LM* curve. Both these results—that an increase in our holdings of foreign assets increases the stock of real balances, and improves the terms of trade—are intuitively plausible.

Next we consider the effects of an increase in the expected rate of depreciation on m and λ. An increase in x reduces the demand for money and shifts the *LM* curve down to *LM″*. In the new equilibrium at E'', the terms of trade have worsened and real balances have fallen. Since $e=\lambda P/P^*$ and both λ and P have risen, the anticipation of depreciation induces an actual depreciation of the exchange rate.

Summarizing this discussion, we can write the equilibrium terms of trade and real balances as follows:

$$\lambda=\lambda(a,x) \qquad \lambda_a<0,\ \lambda_x>0 \tag{11}$$

$$m=m(a,x) \qquad m_a>0,\ m_x<0 \tag{12}$$

Next we use the definition of the terms of trade $\lambda\equiv eP^*/P$ to write:

$$e=\lambda P/P^* \tag{13}$$

or, using the definition of equilibrium real balances, and the equilibrium conditions in the money and goods markets in (11) and (12):

---

[7] The variable k is actually a function of (r^*+x) but we suppress r^* since it is assumed constant.

$$(14)\quad P \equiv M/m(a,x) = eP^*/\lambda(a,x)$$

or

$$(15)\quad e = (M/P^*)\frac{\lambda(a,x)}{m(a,x)}$$

$$= (M/P^*)\,J(a,x); \qquad J_a < 0, J_x > 0$$

We now have an equation for the equilibrium exchange rate in terms of the anticipated rate of depreciation and the stock of external assets. Given the properties of (11) and (13), it is apparent that an increase in external assets leads to an appreciation in the exchange rate. The increase in assets raises real balances which implies a fall in the price level. The rise in wealth in turn raises spending and thus leads to an improvement in the equilibrium terms of trade—a fall in import prices relative to domestic prices. For import prices to fall relative to domestic prices, with domestic prices falling absolutely, the exchange rate must appreciate. An increase in the anticipated rate of depreciation leads to an actual depreciation of the exchange rate, as noted above.

We now move on to discuss the dynamic adjustment of the economy under the assumption of rational expectations. We focus on the dynamics of the exchange rate and the stock of foreign assets. Under the rational expectations hypothesis, in a nonstochastic model, the anticipated rate of depreciation of the exchange rate is equal to the actual rate of depreciation.

To implement formally the assumption of perfect foresight we equate the expected rate of depreciation x, which determines the equilibrium exchange rate in (15) with the actual rate of depreciation $\dot{e}$. Thus setting $x = \dot{e}$ in (15) and inverting that equation to solve for $\dot{e}$ we have

$$(16)\quad \dot{e} = \theta\left(a, \frac{e}{M/P^*}\right) \qquad \theta_1 > 0, \theta_2 > 0$$

It is perhaps worth noting again that (16) describes the dynamic behavior of the exchange rate only under the explicit assumption $x = \dot{e}$, i.e., that the expected and actual rates of depreciation are equal.

The second dynamic equation is that for foreign asset accumulation:

$$(17)\quad \frac{\dot{a}}{r^*} = (1/\lambda)S(m(a,x) + \lambda(a,x)a/r^*)$$

or

$$(17')\quad \dot{a} = \Psi\left(a, \frac{e}{M/P^*}\right) \qquad \Psi_1 < 0, \Psi_2 > 0$$

The derivation of the derivatives of $\Psi(\;)$ is straightforward if tedious; it has to be noted that the condition $\Psi_1 < 0$ is assumed. The assumption that $\Psi_1 < 0$ means that the *total* effect of an increase in external assets on the current account is negative:

$$(18)\quad \Psi_1 = \frac{\partial(\dot{a}/r^*)}{\partial a} + \frac{\partial(\dot{a}/r^*)}{\partial x}\frac{\partial x}{\partial a} < 0$$

The first term is negative since increased assets given expectations raise real wealth and spending, and thus worsen the current account. This was the stability condition in Section I above. The second term involves the effect of increased assets on exchange rate expectations. An increase in assets leads to the expectation of decumulation and depreciation, and therefore reduces real balances and expenditure, thus potentially offsetting the direct effect of higher assets. The stability condition here requires that the former effect dominate and thus is entirely sensible.

Equipped with (16) and (17) we construct Figure 6. The equilibrium is, as is frequently the case with rational expectations, a saddle point; the path the economy follows to the long-run equilibrium at E is shown by the locus FF.[8] Note that the perfect foresight path in Figure 6 shows the conventional

---

[8] There remain some problems in proving that the economy would choose a path such as FF rather than a divergent path. See Fischer and Blanchard for discussion of this point. We choose to study the properties of FF since it is the only convergent path for the exchange rate, given fixed M.

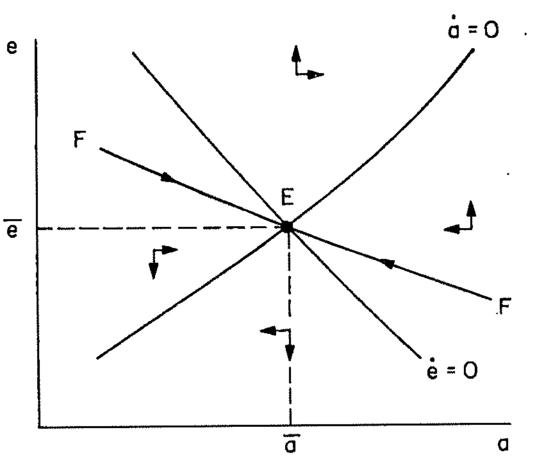


FIGURE 6

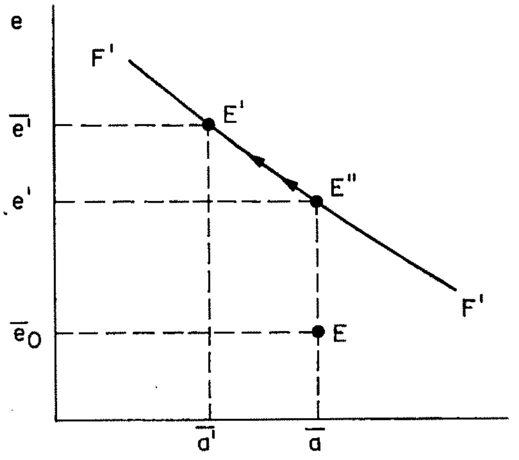


FIGURE 7

association of an appreciating exchange rate and a current account surplus. When assets are above their long-run level, and thus are declining, the exchange rate is below its long-run level, and is thus depreciating.

Consider now the impact effect and adjustment process in response to (unanticipated) disturbances. That question is of interest because it allows a comparison with the case of constant velocity and thus introduces the role of expectations in the determination of short-run exchange rate behavior. We look first at the case of an increase in imports with an offsetting reduction in domestic saving as studied in Figure 3 above. We saw there that the long-run effect is a reduction in external assets and a worsening of the terms of trade. This combination implies a long-run depreciation of the exchange rate since the decline in the value of income from external assets λa reduces equilibrium real money demand and thus implies an increase in prices. For a worsening of the terms of trade the exchange rate accordingly must depreciate.

Figure 7 shows the initial equilibrium at point E, and the new long-run equilibrium, with lower $\bar{a}$ and higher $\bar{e}$, at E'. Before the unanticipated increase in imports, the exchange rate is $\bar{e}_0$ and the stock of foreign assets $\bar{a}$. In the short run, external assets are given at $\bar{a}$. But the reduced saving and increased import spending create the expectation of depreciation, reducing the demand for real balances, increasing the price level, and reducing real spending (through a real balance effect in the goods market). The terms of trade deteriorate to maintain goods market equilibrium; the deterioration of the terms of trade imply an immediate adjustment of the exchange rate to E''. The impact effect of the disturbance is to cause a one-time discrete depreciation; thereafter the exchange rate depreciates along the path to E' as assets are decumulated.

In discussing the case of an unanticipated increase in exports, studied in Figure 2 above, we sketch the use of the graphical apparatus set out in this section. Figure 8 shows the initial equilibrium at E. We know from Figure 2 that in the new long-run equilibrium, a is higher than it was at E, and λ is lower. We know further that λa is unchanged between the two equilibria (see Section I). Since y is constant, and since $x=0$ in both long-run equilibria, the stock of real balances is unchanged and the price level is unchanged. Since λ falls and P is unchanged, the exchange rate must appreciate between old and new equilibria. We therefore know that the new equilibrium in Figure 8 is to the southeast of E; it is shown at E'. Passing through E' are the $\dot{e}'=0$ and $\dot{a}'=0$ loci.[9]

---

[9] It is possible, though complicated, to work through all the equations set out in this section to reach the conclusion that E' is to the southeast of E. It is much easier to use the properties of the long-run equilibrium.

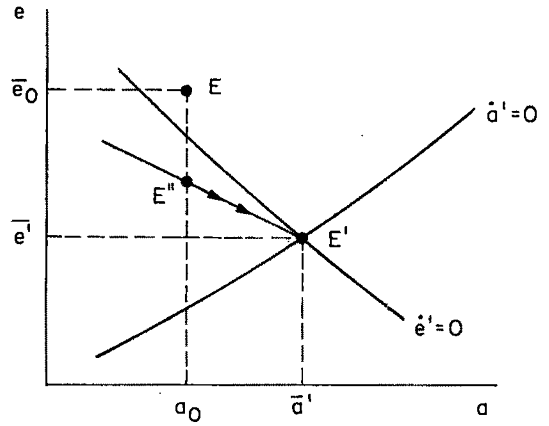


FIGURE 8

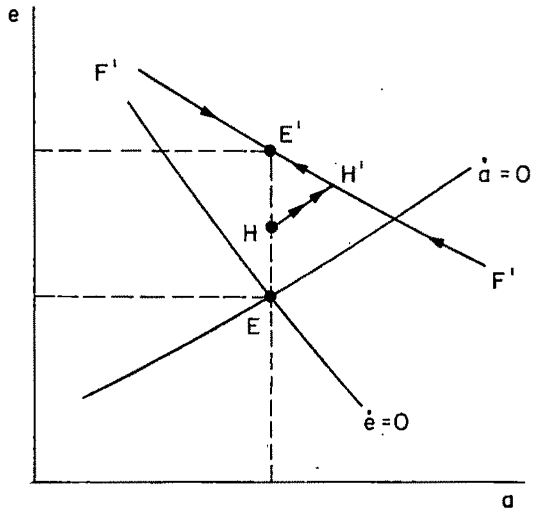


FIGURE 9

The only ambiguity that apparently remains is whether the exchange rate initially appreciates (as in Figure 8) or depreciates. However, there is no ambiguity and Figure 8 is correct. This can be shown formally by proving that the ($\dot{e}'=0$) locus shifts downward. More intuitively, we know that the exchange rate will be appreciating on the path to E'. Even were x to remain equal to zero, we would have λ fall as in Figure 2; now we have higher real balances, therefore more wealth, and a larger fall in λ is required. Since P and λ both fall, e has to fall. Again, the transition path is one on which an appreciating exchange rate is accompanied by a current account surplus.

We may also ask how the inclusion of exchange rate expectations in this section affects the dynamics as compared with Section I, where velocity was assumed constant and expectations were ignored. Allowing for expectations increases the magnitude of the initial decline in the terms of trade. Anticipated appreciation increases the demand for money (and therefore the stock of real balances, wealth, and the demand for our goods) and thus requires a larger initial fall in the terms of trade.

No exchange rate paper can be complete without examining the effects of a change in the stock of money. It is easy to show that an unanticipated increase in the stock of money merely increases the exchange rate proportionally and immediately, and that there is no accompanying process of accumulation. The immediate adjustment is, of course, a consequence of the assumption of continuous full employment and price flexibility.

III. Anticipated Disturbances

Up to this stage, the association between an appreciating exchange rate and a current account surplus has been strong. However, this connection is modified when we analyze the effects of anticipated disturbances.[10]

We study now the adjustment process to a current announcement of a *future* increase in the nominal quantity of money. We immediately note that at the point in time where the money stock actually increases, we are back in the framework of analysis of the previous section shown as the path $F'F'$ in Figure 9, with the new long-run equilibrium at point E'. What happens, though, in the period between the announcement and the actual increase in money?

In the transition period, the adjustment process is governed by equations (15), (16), and (17′). In terms of Figure 9 this means that the initial $\dot{e}=0$ and $\dot{a}=0$ loci remain

---

[10] Anticipated future disturbances and their impact on exchange rates have been studied by Wilson.

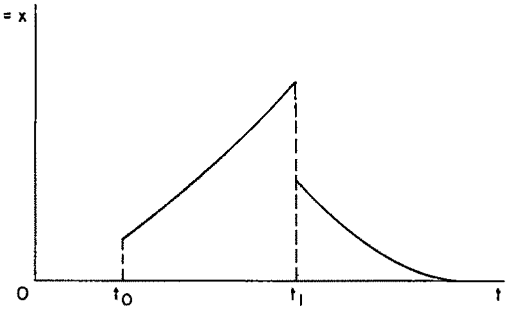


FIGURE 10

relevant for the analysis. We have also to observe that any jump in the exchange rate has to occur immediately and cannot lie further ahead along the adjustment path. This must be true since otherwise there would be the anticipation of infinite capital gains and therefore incipient capital outflows that would immediately move the rate. With these two considerations in mind, we observe from Figure 9 the qualitative properties of the adjustment path.

There will be an immediate depreciation of the exchange rate at the time of announcement of a future increase in the money stock. From then on, the exchange rate will continue to depreciate and the anticipation of depreciation, by lowering real balances, wealth and spending, will give rise to a current account surplus. The process continues until we reach a point like H' exactly at the time the money supply increases. Exchange rate depreciation continues, but the increased real balances, wealth, and spending now lead to a deficit and decumulation of external assets until the initial real equilibrium is reattained.

We can gain further insight into the adjustment process by studying the behavior of the rate of depreciation shown in Figure 10. The rate of depreciation, $\dot{e} = x$, is described in (16). It is apparent that at the time of announcement t_o, the only change is the jump in the level of the exchange rate, and thus we have the anticipation of a depreciating rate. Along the adjustment path from H to H' in Figure 9, assets are growing as is the exchange rate. Accordingly, using (16), the rate of depreciation is increasing.

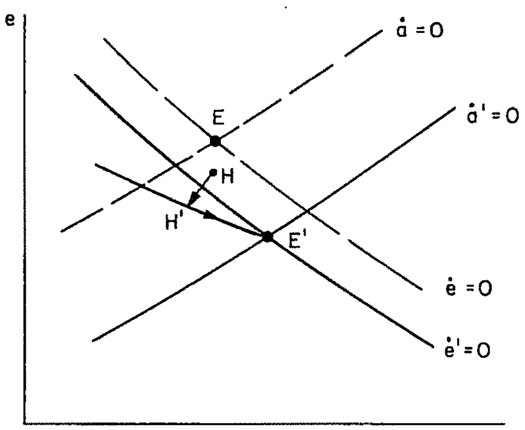


FIGURE 11

Once point H' is reached at time t_1, there is a discrete reduction in the rate of depreciation. Since the exchange rate cannot jump now, but nominal money does increase, the rate of depreciation must decline. From then on, depreciation continually decreases until it reaches zero in the steady state.

The adjustment path of the current account and of the rate of depreciation of course qualify, in a very important way, the popular association between monetary disturbances, depreciation, and the external balance. The striking results here are that the anticipation of a money supply increase should initially lead to a surplus, and that the rate of depreciation should decrease exactly at the point in time where the increase in money actually occurs.

These results operate through the effects of the anticipated change in the money stock on absorption. The anticipated increase in the money stock causes an anticipated—and therefore actual—depreciation of the currency, reducing the demand for money, and thereby reducing real balances. Savings are thus increased, and assets are imported through the current account surplus. After the money stock is actually increased, there is a real balance effect working in the opposite direction, causing dissaving through the current account.

To confirm that the incorporation of anticipated policy changes generally changes the character of the results of Section III, we will take up our running example of an anticipated increase in exports. Figure 11 contains the diagram; it is to be compared with Figure 8.

The appearance of Figure 11 is very similar to that of Figure 8. This time the exchange rate appreciates initially and continues appreciating as assets are decumulated in anticipation of the increase in exports. The initial exchange rate is chosen such that the economy reaches point H' at the moment the increase in exports occurs. The economic explanation of the path followed is similar to that outlined above; the anticipated appreciation may be thought to be the driving force in the dynamics.

IV. Concluding Remarks

Recent models of exchange rate determination and dynamics have emphasized asset stock equilibrium as the relevant framework. The study of exchange rate dynamics has typically been formulated in a macro-economic setting that emphasizes the speeds of adjustment of goods and assets markets, and in which expectations dominate the short-run behavior of the exchange rate.

This paper focuses instead on the association between the current account and the exchange rate. It retains many elements of the recent approaches including asset equilibrium, and the distinction between anticipated and unanticipated disturbances to the economy. But it extends recent approaches by giving an important role to the accumulation of assets over time through the current account. Asset accumulation, together with expectations, dominates the dynamic behavior of the exchange rate.

One point of difference with asset market models of the exchange rate concerns the question of overshooting. Monetary changes in models where assets markets move rapidly relative to goods markets lead to a more than proportionate depreciation of the exchange rate. The reason is that real balances increase because of the predetermined price level, that equilibrium nominal interest rates decline, and that international interest arbitrage requires an offsetting expectation of appreciation. The only way, with the anticipation of long-run depreciation, that we can have an expectation of appreciation is for the exchange rate to overshoot. The model differs in important respects from our presentation. Here the price level is fully flexible so that (current) monetary disturbances have no real effects. Open-market operations, a monetary expansion with a contraction in external assets, lead to a decline in real balances and therefore to an appreciation that in the short-run is more than proportional to the fall in money. Thus there is overshooting in the present model, not because prices are sticky, but because in the short-run external assets are predetermined.

Our model has the important shortcomings of a very restricted menu of assets, and of short-run price flexibility and continuous full employment. The compensation is the introduction of the current account as an important element in open-economy macroeconomics and exchange rate determination. In doing so, the paper provides a theoretical rationale for the popular view that there is an association between the current account and the behavior of the exchange rate. However, such a view has to be qualified because anticipated disturbances that will ultimately depreciate the exchange rate can, initially and in contrast to the conventional view, lead to the combination of a current account deficit and appreciating exchange rate.

REFERENCES

O. Blanchard, "Backward and Forward Solutions for Economies with Rational Expectations," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 114–18.

R. Dornbusch, "A Portfolio Balance Model of the Open Economy," *J. Monet. Econ.*, Jan. 1975, *1*, 3–20.

———, "Capital Mobility, Flexible Exchange Rates and Macroeconomic Equilibrium," in Emil Claassen and Pascal Salin, eds., *Recent Issues In International Monetary Economics*, New York 1976.

S. Fischer, "Anticipations and the Non-Neutrality of Money," *J. Polit. Econ*, Apr.

1979, *87*, 225–52.

J. M. Fleming, "Domestic Financial Policy under Fixed and Floating Exchange Rates," *Int. Monet. Fund Staff Papers,* Nov. 1962, *9*, 369–79.

R. Flood, "A Model of Asset Trade," unpublished manuscript, Univ. Virginia 1977.

J. A. Frenkel and C. A. Rodriguez, "Portfolio Equilibrium and the Balance of Payments: A Monetary Approach," *Amer. Econ. Rev.*, Sept. 1975, *66*, 674–88.

P. J. K. Kouri, "Essays on the Theory of Flexible Exchange Rates," unpublished doctoral dissertation, M.I.T. 1975.

______, "The Exchange Rate and the Balance of Payments in the Short Run and in the Long Run: A Monetary Approach," *Scand. J. Econ.*, No. 2, 1976, *78*, 280–304.

F. Modigliani, "The Life Cycle Hypothesis of Saving Twenty Years Later," in Michael Parkin and A. R. Nobay, eds., *Contemporary Issues in Economics:...Warwick, 1973*, Manchester 1975.

Robert A. Mundell, *International Economics*, New York 1968.

C. A. Rodriguez, "The Role of Trade Flows in Exchange Rate Determination: A Rational Expectations Approach," unpublished manuscript, Columbia Univ. 1977.

C. A. Wilson, "Anticipated Shocks and Exchange Rate Dynamics," *J. Polit. Econ.*, June 1979, *87*, 639–47.

The Short-Run Relation between Growth and Inflation in Latin America: A Quasi-Rational or Consistent Expectations Approach

By JAMES A. HANSON*

This paper explores the short-run relation between inflation and growth in Latin America. The topic was the subject of heated debate between structuralists and monetarists in the 1960's, but extensive empirical exploration has only begun.[1] The evidence presented here should provide a useful complement to the numerous studies of the relation between output, employment, and inflation in developed countries. Latin America provides a broad set of well-documented experiments with which to test hypotheses regarding inflation, far broader in fact than the Group of Ten.

To examine the relation between growth and inflation, I use a variant of the Phillips models formulated by Edmund Phelps, Robert Lucas and Leonard Rapping, and Lucas (1973). Deviations from normal output or growth rates are attributed to the difference between a reduced form of actual and expected prices, the latter being characterized by rationality, as defined below. The major conclusions obtained using this model are:

1. In Brazil, Chile before Allende, Colombia, Mexico and Peru, a small though significant relation exists between output or growth and "unexpected" inflation. This result is substantially stronger than those obtained in Robert Barro's (1974, 1979) studies of Latin American inflations.

2. Ten percentage points of unexpected inflation produce about one extra percentage point of growth. This effect is statistically indistinguishable across the five countries once the different processes of monetary supply and corresponding formulation of expectations are taken into account. This result strongly supports Lucas' hypothesis that the reaction to unexpected inflation depends on the predictability of inflation, because predictability in the five cases is seen to be similar, once one allows for differences in monetary supply processes.[2] This effect is also significantly smaller than those obtained by Lucas (1973) for the developed countries with more predictable inflations, confirming his hypothesis of a near vertical Phillips curve when nominal demands are unpredictable.

To put it another way, this paper may be considered as an attempt to add a "third data point"—moderately inflationary countries such as Colombia, Peru and Mexico—to what Lucas (1973) refers to as "two data points"—the developed countries and the inflationary countries of Argentina and Paraguay.[3] In addition, the cases of Brazil

---

*Brown University. This paper was completed while I was a visiting scholar at the International Finance Division of the Board of Governors of the Federal Reserve System. I am grateful for the use of computer facilities at *CEPAL*, Brown University, and the Board of Governors, and for comments from Roque Fernandez, Robert Gregory, Herschel Grossman, Dale Henderson, Robert King and Jerome Stein, though I alone bear full responsibility for the study. The analysis and conclusions of this paper should not be interpreted as representing the views of the Board of Governors of the Federal Reserve System or members of its staff. More extensive statistical results are contained in my *International Finance Discussion Paper*.

[1] For a review of existing work, see the paper by Roque Fernandez and myself. Summaries of the structuralist view may be found in Roberto Campos, Enrique Sierra, and Susan Wachter, among others.

[2] Lucas uses the variance of detrended nominal demand as a measure of its predictability because he assumes that variable can be characterized by a random walk with trend. With a more complicated money supply process, predictability of inflation becomes the key variable in the reaction to unexpected inflation.

[3] It is worth noting that there is a slight inverse relationship between the coefficient of unexpected inflation and the variability of inflation within Lucas' developed country "data point," tending to confirm his hypothesis. Both within Lucas' (1973) sample and in Latin America, as discussed in Fernandez and myself,

and Chile add weight to Lucas' "second data point." The results suggest that there may be little difference between the predictability of high and moderate inflations. Therefore, the observed, low but similar, response to unexpected inflation strongly supports Lucas' hypothesis.

3. These results are in terms of growth rates. Lucas' (1973) formulation hypothesized that the level of output is a function of unexpected inflation, lagged output, and a time trend. However, in the five countries studied, there is no statistical distinction between the formulation in growth rates and that in levels of output as a (reduced-form) function of unexpected inflation.[4]

The growth form implies that increases in growth achieved through changes in prices last only as long as the price change is unexpected. Any temporary rise in the real growth rate and output above the normal values, due to unexpected inflation, will be "paid for" as expected inflation overshoots actual. One implication of the growth form is that if, following unexpected inflation, expected inflation remained equal to actual, then output would remain above normal. Such a result can be justified in many ways, for example, through the capacity effects of a one-shot increase in money or the persistent effects of supply shocks, terms of trade changes, or inappropriate exchange rates. Thus the result should not be surprising, although a one-shot change in unexpected inflation is an unlikely hypothesis regarding expectations or the politics of money supply. More to the point, the formation of expectations in the paper is assumed to and does predict inflation on average over the sample, generally with serially uncorrelated errors. In this sense expectations are rational, though perhaps consistent would be a better term. Neither persistent unexpected inflation nor a price level persistently above trend are permitted. As a result, serially correlated deviations from normal growth do not occur in the countries studied, although deviations from trend output are persistent. Moreover, the independent influence of unexpected changes in inflation on the persistent deviations from trend output is quite small.

4. It is difficult to establish the particular process of money growth and expectation formation, even in the rapid, erratic inflations studied. In the two rapid inflations, lagged values of inflation seem to work better in determining output growth than the monetary growth variables used by Lucas (1973) and Barro (1974, 1977, 1979). A possible explanation is that in these cases lagged inflation is at least as good a predictor of monetary growth as lagged monetary growth, perhaps because government policy and the balance of payments are affected by lagged inflation, which is only imperfectly correlated with past monetary growth. Thomas Sargent and Neil Wallace have provided an alternative justification for the use of lagged inflation as a rational predictor in hyperinflationary situations.

It also appears that in Latin American economies with more rapid and variable inflation, more "up-to-date" data are useful in predicting inflation. However, this fact may simply reflect the peculiar seasonality of price changes and of the money supply process in these countries.

Another interesting point is that the results for Brazil were substantially improved by dividing the sample into pre- and post-1965 monetary regimes, which were statistically distinguishable in terms of the monetary growth process and the corresponding formation of expectations. This result may indicate that the process of forming expectations changes rapidly to correspond to a new monetary regime. However, the hypothesis could not be tested with data from the other countries owing to the shortness of objectively identifiable, monetary regimes.

5. Lagged unexpected money growth (and inflation) generally had an insignificant effect on current output, especially when the previous year's output was taken into account. Estimates which omit the previous year's output display autocorrelated

---

there is a nearly proportional relation between inflation and its standard deviation, which makes it difficult to distinguish Lucas' hypothesis from competing ones.

[4] Inspection indicates this is also true in about 60 percent of the cases studied by Lucas (1973).

residuals. These findings differ from Barro's *U.S.* (1977) and Mexican (1979) studies and seem to indicate the persistence of relatively strong, nonmonetary disturbances. Perhaps such disturbances are related to supply shifts or to the simple treatment of the foreign sector in this paper, hypotheses which will be explored in later work.

Section I outlines the basic model which is used for empirical estimation and discusses the expectations processes which are used. It also discusses the growth equation formulation. Section II summarizes the empirical results on the money supply process and the determination of output when the expectations process is assumed to be based on a prediction of monetary growth. Section III summarizes the results when lagged inflation is used to predict inflation. Section IV discusses the results obtained when two different money supply processes are hypothesized for Brazil, and Section V compares the relative importance in output determination of lagged effects of unexpected inflation and of supply shocks, proxied by lagged output. Section VI very briefly summarizes the results.

I. The Basic Model

This paper uses a model of the aggregate supply relation between output and inflation based on work of Phelps, Lucas and Rapping, and Lucas (1973). Suppliers increase (decrease) output above (below) its normal trend when prices turn out better (worse) than expected. This argument is often applied to labor supply; when wages are higher than expected, above normal amounts of labor are supplied, and output increases faster than trend, because workers seek to take advantage of a good wage. It might be especially applicable to Latin American coun tries, where the existence of secondary workers and rural-urban migration, as well as poor information networks, may create some short-run money illusion in the labor supply function.

Lucas (1973) writes the aggregate supply function as

$$(1)\qquad Y_t = a_0 + a_1\left(P_t - P_t^E\right) + a_2\,(\text{time}) + a_3 Y_{t-1} + u_{t1}$$

where

$Y_t = log$ of aggregate supply at time t
$P_t = log$ of price level at time t
$P_t^E = E(log$ of $P_t|$Information available at time t)[5]
$a_1 > 0,\ a_3 \geqslant 1$

There are two problems in estimating (1): specification of an aggregate demand equation which, together with (1) and P^E, will simultaneously determine output and prices, and specification of P^E.

Rather than explain the behavior of aggregate demand subject to some income or wealth constraint, this paper simply applies the constraint directly. In particular, it is assumed that monetary disturbances do not spill over into the world economy because of variations in the effective exchange rate and that households are satisfied with their nominal money holdings in every period, vis-à-vis the levels of prices, income, wealth, and opportunity costs of money holdings. That is, demand for output exceeds supply until prices, incomes, and/or the opportunity cost of money adjust to yield

$$(2)\qquad M_t^S = M_t^D \equiv P_t + b_0 + b_1 Y_t + b_2 r_t + u_{t2}$$

where $M_t^i = log$ of money stock supplied (S) or demanded (D) and $r_t =$ the *log* of the relevant opportunity cost of holding money at time t, which is omitted in the empirical work owing to its statistical insignificance.[6] Complete stock adjustment may seem a strong restriction, but the lack of autocorrelation in the empirical results suggest it is a reasonable approximation. The approach also seems to neglect fiscal policy, but again it may not be a bad approximation for these economies, where monetary and fiscal policy have been closely associated. (See Arnold Harberger, 1975, and Werner Baer and Paul Beckerman.)

---

[5] Lucas (1973) derives (1) from an explicit aggregation of micro behavior, not shown here. Some difficulties in statistical estimation may arise when national accounts, which are Paasche or Laspeyres weighted, are substituted for Lucas' geometrically weighted outputs. Lucas (1975) gives some motivation for inclusion of lagged output.

[6] Results are available on request. Money demand could be written as a function of the weighted average of prices and expected prices without changing any of the results.

Lucas (1973) hypothesized that expectations are formed rationally in the sense of John Muth, that is, using all relevant information as if the relevant economic theory for its application were known. In the present context this would mean that expectations take into account the output-price adjustment process described by (1) and (2). In addition, as Lucas (1970), among others, has pointed out, if the government systematically varied aggregate demand in response to indicators, then such behavior would only have real effects until it was incorporated into the public's forecasts.

In hypothesizing an expected rate of inflation in the economies under consideration, where electronic computing facilities were and are limited, and where markets, knowledge of economic theory, and information are imperfect, it seems appropriate to use a simplified approach to expectation formation, incorporating the role of government. One general approach which is used here follows Lucas (1973) by assuming (the *log* of) money is a random walk with trend (m), and solving (1) and (2) to obtain

$$Y_t = c_1 + a_1 DM_t + a_2 \text{ time} + a_3 Y_{t-1} + u_{t3} \tag{3}$$

with m becoming part of the estimated constant and DM_t the independent variable. (The D notation indicates a time derivative.) That is, (3) is a reduced form using the hypothesized money supply process and the stock money demand equation (2).[7] Note that it implies lagged values of money have negative coefficients.

In a slightly more complicated model, Barro (1977) hypothesizes that money growth depends on, and therefore, is rationally predicted by, previous values of money growth and other variables such as government spending, the unemployment rate, and the balance of payments. The difference between actual and expected prices is a linear function of the residuals (DMR_t) from an estimated equation of the form

$$DM_t = d_0 + d_1 DM_{t-1} + d_2 DM_{t-2} + d_3 DM_{t-3} + u_{t4} \tag{4}$$

where in this paper the other possible determinants of DM_t have been omitted owing to lack of data. The usual stability conditions may be tested on the d_i; negative coefficients might indicate corrective policies. The analogous form to (3) is

$$Y_t = c_2 + a_1 DMR_t + a_2 \text{ time} + a_3 Y_{t-1} + u_{t5} \tag{5}$$

with the ith lagged monetary growth rate having the coefficient $-a_1 d_i$.

The second general approach uses past inflation as a predictor. While this approach sometimes has been considered *ad hoc*, Sargent and Wallace, among others, have shown that it can be rational, in particular, when government "is financing a roughly fixed rate of real expenditures by money creation" (p. 336). Dean Dutton obtains some success in explaining Argentine quarterly data with a model of constant real expenditures, real taxes which are inversely related to inflation, and a monetized deficit. Harberger (1975), Baer and Beckerman, and Ricardo Ffrench-Davis have all indicated the importance of monetary emission as a source of Latin American government revenues. Thus, it seems important to explore this process of forming expectations in the case of rapid Latin American inflations.

As a simple way of examining this hypothesis, lagged inflation rates were substituted for the lagged monetary growth rates in (4), with the residuals substituted for DMR in (5). Thus money growth is still being predicted. A second alternative was also investigated, in which inflation was assumed to be predicted directly by past inflation, as in the well-known adaptive expectations hypothesis used by Philip Cagan. But if it is inflation rather than monetary growth which is predicted, then it seems reasonable to assume that households forecast next period's price level by taking (the *log* of) past prices as given and adding to it their prediction of inflation—$P_t^E = P_{t-1} + DP_t^E$. The actual price level P_t equals the previous level plus the solution for price changes taken from a difference version of (2), $P_t = P_{t-1} + DP_t = P_{t-1} + DM_t - b_1 DY_t + Du_{t2}$. Substituting into (1) yields

---

[7] The actual method of solution is substantially more complex and involves questions of stability and convergence. For details, see John Muth, Lucas (1970), and Barro and Stanley Fischer.

$$(6) \quad Y_t = a_0 + \frac{a_1}{1+b_1a_1}(DM_t - DP_t^E) + \frac{a_2}{1+b_1a_1}\,(\text{time}) + \frac{a_3+b_1a_1}{1+b_1a_1}\,Y_{t-1} + u_{t6}$$

a reduced form roughly analogous to (3) and (5).

Two specific versions of direct inflation prediction were tried. The first simply assumes adaptive expectations with full adjustment to errors $DP_t^E = DP_{t-1}$ yielding

$$(7) \quad Y_t = a_0^* + a_1^*(DM_t - DP_{t-1}) + a_2^*\,(\text{time}) + a_3^* Y_{t-1} + u_{t7}$$

Though this might seem a strong restriction, it has the merit of simplicity and provides a useful standard of comparison for more complex models of expectations. It is also an approximation familiar to students of Latin American inflation. (See Harberger, 1963, Robert Vogel, Edmund Sheehy, and Wachter.)

A more complicated version permits inflationary expectations to be determined by the history of past rates of inflation, plus a constant:

$$(8) \quad DP_t = d_1^* DP_{t-1} + d_2^* DP_{t-2} + d_3^* DP_{t-3} + d_0^* + u_{t8}$$

yielding

$$(9) \quad Y_t = a_0^* + a_1^*(DM_t - DP_t^E) + a_2^*\,(\text{time}) + a_3^* Y_{t-1} + u_{t9}$$

where the d_i^* are estimated without restriction. The adaptive expectations model implies $\Sigma d_i^* = 1$; $d_{i+1}^* = (1 - d_1^*)d_i^*$, $i \geqslant 1$, $d_0^* = 0$. However, one can easily imagine some cyclical behavior of world prices or systematic government policy in response to abnormal price trends which yields some negative coefficients.[8] If $d_i^* = 0$, $i > 0$, then inflation is a random walk with trend.

The monetary growth and the inflation prediction approaches may be thought of as alternative hypotheses regarding the formulation of monetary growth policy and its optimal prediction.[9] Equations (3) and (5), when using past monetary growth to generate *DMR*, are based on the assumption that policy depends on past monetary growth. In contrast equation (5), when using past inflation to generate *DMR*, and equations (7) and (9), can be thought of as assuming a monetary growth policy based on past inflation. The two approaches would be equivalent if policy and the money demand function were nonstochastic and there were no errors in measurement. In that case, *P* and *M* are exact functions of each other and equations (4) and (8) are equivalent. Given the stochastic nature of the money demand function and of policy, and assuming the stochastic element is serially uncorrelated (a proposition to be examined empirically below), then the relative quality of the two predictions, in the sense of a smaller forecast error for the inflation rate, depends upon the actual policy rule which is used.[10]

---

[8] Adolfo Diz rationalizes such coefficients in Argentina as overreaction to monetary expansion, and Sheehy reports a significantly negative coefficient for twice-lagged money in his regressions explaining Argentine inflation.

[9] The assumption of complete stock adjustment of money balances means history is useful only insofar as it conveys information about the future path of monetary policy; there being no "overhang" of monetary balances which must be absorbed or no lagged effects in subsequent periods. However, an alternative interpretation would be that (3) and (5) are constrained versions of the lagged impact of money. As Barro (1977) points out, it is difficult to separate such lagged influences from the influence of unexpected money unless money is predicted with something other than lagged values. Nonetheless one might suspect that policy reactions dominate, rather than the lagged impacts of money, if unconstrained coefficients of lagged values appear with negative signs, and if lagged money growth is a relatively poor predictor of inflation.

[10] This point neglects measurement error of *M*, which plays an important role in Sargent and Wallace's analysis.

That is, lagged inflation may provide a better predictor if past stochastic elements in the difference version of (2) affect M_t.[11] This may well be true—imagine the plight of a finance minister who justifies his inaction on inflation by reference to a random negative shift in money demand.

Christopher Sims' tests of timing also are used in this paper, in conjunction with (8) and variants of (4), to establish the policy rule. Wachter's application of the tests to quarterly data from Brazil, Chile, and Mexico indicates one cannot reject the hypothesis that past inflation influences money creation, though the result does not appear in annual data. Fernandez' ARIMA results, also with quarterly data, indicate that lagged values of prices account for most of the current variation in prices.

If past inflation does influence monetary creation, then lagged rates of inflation or variables correlated with them could reduce the significance of present and lagged rates of monetary creation in estimated equations of inflation.[12] This might explain the significance of structural variables such as lagged wages in the inflation regressions of Harberger (1963), Carlos Diaz Alejandro (1970, 1972) and Sheehy. While equations such as (8) seem to imply that past inflation causes present inflation, this is only because of the rule for monetary growth. A new monetary rule would change inflation in a way that would be unpredictable using equations (7) and (8), but eventually predictable using the money demand function (2).

It should be noted that all of the previously discussed forms imply use of information from the future at each point in time. This is because the estimated trend and the coefficients of lagged values which are hypothetically used for prediction are based on the whole sample, not just information previous to each observation, as is the case in adaptive expectations. This is perhaps less worrisome than might appear. Estimated expectations of inflation are based on a simplistic form which obviously omits many variables that might aid in prediction and with a correspondingly larger forecasting error than actual expectations. The use of an estimated constant (and coefficients) for the sample requires the predictor to be correct on average and perhaps more closely approximate actual expectations.

Finally, since interest has focused on the relation between growth and inflation, and since it is natural to think in a growth framework when dealing with developing countries, versions of equations (3), (5), (7), and (9) have also been estimated in (logarithmic) growth rates. However, rather than simply taking first differences, estimates were made under the constraint $a_2=0$, $a_3=1$.

$$DY_t = a_1 DM_t + a_2' \tag{3'}$$

$$DY_t = a_1 DMR_t + a_2' \tag{5'}$$

$$DY_t = a_1^*(DM_t - DP_{t-1}) + a_2^{*\prime} \tag{7'}$$

$$DY_t = a_1^*(DM_t - DP_t^E) + a_2^{*\prime} \tag{9'}$$

These equations have the straightforward interpretation that supply normally grows at some average or "natural" rate, while actual output grows by less or more due to unexpected inflation.[13] If they are statistically indistinguishable from the corresponding level equations, then the simple first-difference form also cannot be rejected, as (3')–(9') are essentially first-difference equations with additional constraints.[14]

Acceptance of these equations implies that output could persistently differ from normal levels for long periods, owing to the large coefficient of lagged output. Literally, a single disturbance or an unexpected rise in M or DM in (3′), or DMR in (5′), would keep income above its "normal" level forever. In (7′) or (9′) output would return to normal after a single disturbance in M, though not after one in DP. Such an empirical result can be explained in many ways. Supply shocks, abnormal terms of trade, and incorrect exchange rates may persist for long periods, thus affecting output levels. The usual *IS-LM* model would imply that a single disturbance in M would raise the capital stock with output effects that would persist well beyond most small samples' limits; many other models would have similar implications for capital accumulation. It is therefore not surprising that the coefficients of lagged output are not statistically different from one in 60 percent of the countries studied by Lucas (1973). Moreover, use of the level form biases the coefficient of lagged output downward, and tends to produce autocorrelated residuals if a_3 is close to one; this is true to an even greater degree if detrended data are used. Thus the appropriate estimating form is not obvious a priori.

---

[11] Marcelle Arak's critique of Lucas' (1973) test suffers from her neglect of the stochastic nature of the nominal demand function.

[12] Robert King argues rational expectations might cause prices to "lead" money.

[13] Notice that in estimating (7') and (9'), a_1^* multiplied by the average difference between DM_t and DP_{t-1} or DP_t^E becomes a part of the constant term, but that product is simply the (estimated) income elasticity of demand for money, multiplied by the average growth of income and a_1^*, so the interpretation of $a_2^{*\prime}$ is similar to a_2^*.

[14] They are not quite constrained versions of the first-difference equations. While a_3 is constrained to equal one, the coefficient of the lagged unexpected inflation variable is constrained to zero, rather than to a_1. The results of Section V show this difference is not important.

While the results of this paper support the applicability of the growth rate form, they also indicate that the possibility of raising output through a single increase in unexpected inflation remains theoretical. First, such an increase would produce autocorrelated errors in predicting inflation in a small sample, but the prediction equations in this paper seem to display serially uncorrelated residuals, though the test is weak. It is not surprising that this particular monetary experiment was not performed; if it had occurred then its effects probably would have stimulated further experiments, with still further reactions by the public and then the policy makers, thereby destroying the initial experiment's uniqueness. Second, the lagged independent effects of unexpected monetary growth on output are statistically insignificant once lagged output is introduced as an explanatory variable, implying a tenuous relation between unexpected inflation and capacity growth.

Previous work by Lucas (1973) suggests that a_1 is inversely related to the variability of inflation. His estimates for fourteen low-inflation countries generally yield values between .8 and .4. While there is a slight inverse relation between a_1 and the variability of inflation in these countries, most of the support for his view comes from his "second" data point, the two outlying Latin American countries—Argentina and Paraguay—where coefficients are below .03 and insignificant. Barro (1979) finds similarly low insignificant coefficients for Brazil and Colombia, but his coefficients for Mexico are jointly significant and their sum is in the range of low-inflation countries.

II. Empirical Results using Unexpected Changes in Monetary Growth to Explain Output

Table 1 presents estimates and statistics of fit for equations (3) and (3′)—expectations which assume money is a random walk with trend. The growth rate formulation is statistically indistinguishable from Lucas' formulation in terms of levels; the increase in sum of squared residuals (*SSR*) due to the two growth rate restrictions (compare upper and lower portions of Table 1) yields an F-statistic of less than 2 for all countries except Brazil (4.1). Operationally, this means the positive impact of a one-shot increase in money or monetary growth cannot be discounted.

Unexpected monetary growth of ten percentage points raises output about one percentage point above the trend in the countries with moderate inflation—Colombia, Mexico, and Peru. However, in the level formulation this coefficient is significantly positive only for Mexico; in the growth formulation only for Mexico and Colombia. The Durbin-Watson statistic indicates no autocorrelation for the growth form for Mexico and Peru and only marginal autocorrelation for Columbia, providing support for the approximation of complete stock adjustment within one period.[15]

---

[15] Notice, however, that money is defined as an annual average of end-of-month figures to smooth the data and avoid the erratic year-end bulge in money which may have contributed to Barro's (1979) poor results. It also tends to bias the results toward accep-

TABLE 1—OUTPUT AND GROWTH OF OUTPUT AS FUNCTIONS OF MONEY GROWTH

	Brazil 1952–74	Chile 1952–70	Colombia 1951–74	Mexico 1950–74	Peru 1951–74	Pooled[d]
			$(3) Y_t = a_0 + a_1 DM_t + a_2 \text{(time)} + a_3 Y_{t-1}$			
a_0	−.074	−.246	−.286	−1.11	−.487	–
	(.2)	(.7)	(.5)	(1.9)	(1.2)	
a_1	−.130[a]	−.026	.092	.133[a]	.092	−.018
	(2.8)	(.4)	(1.8)	(2.3)	(1.2)	(0.7)
a_2	.00086	.0048	.0028	.016	.014	–
	(.1)	(.9)	(.5)	(2.0)	(1.5)	
a_3	1.02[a]	.88[a]	.95[a]	.74[a]	.71[a]	–
	(9.3)	(5.6)	(8.1)	(5.6)	(4.1)	
DF	19	15	20	21	20	99
SSR	.0119	.0054	.0036	.0051	.0116	.0458
SEE	.0250	.0189	.0134	.0156	.0242	.0216
R^2	.997	.993	.999	.999	.996	.99
DW	2.1	1.24[b]	1.45[b]	1.82	1.42[b]	–
			$(3') D\dot{Y}_t = a_2' + a_1 DM_t$			
a_2'	.095[a]	.051[a]	.032[a]	.047[a]	.036[a]	[e]
	(5.8)	(3.1)	(4.3)	(6.0)	(3.2)	
a_1	−.082	−.050	.11[a]	.134[a]	.100	.004
	(1.7)	(1.0)	(2.6)	(2.2)	(1.5)	(0.1)
SSR	.0167	.0057	.0038	.0061	.0139	.0536
SEE	.0282	.0184	.0132	.0163	.0252	.0223
R^2	.12	.06	.23	.18	.09	.21
DW	1.36[c]	1.30[c]	1.38[c]	1.97	1.58	–
SD of *DM*	.1244	.0857	.0624	.0559	.0784	–

Sources: IFS, Ffrench-Davis (Chile), U.N. *Yearbook of National Income Accounts Statistics* (Peru 1971-74). Data available on request. For more details see my earlier paper.

Notes: Absolute value of *t*-statistic in parenthesis. *SD*=standard deviation, *DF*=degrees of freedom, *SSR*=sum of squared residuals, *SEE*=standard error of estimate, *DW*=Durbin-Watson statistic.

[a] Significant at 95 percent level.

[b] Possible autocorrelation at the .01 level.

[c] Possible autocorrelation at the .025 level.

[d] Respective individual values for a_2 (and *t*) are .0018 (.3), .0051 (.9), .0015 (.2), .0165 (1.5), .0199 (2.6); for a_3 (and *t*): 1.00 (10.6), .87 (5.5), .99 (5.3), .74 (4.1), .61 (4.1). A pooled regression, weighted by the inverse of individual country *SEE*s, yielded an *SSR* which exceeded the weighted total of the individual *SSR*s by the significant amount of 20.2 percent.

[e] Individual country constants (and *t*) were: .058 (5.8), .023 (2.5), .045 (9.2), .047 (7.2), .058 (12.9). The corresponding weighted pooled regression increased the weighted total of the individual *SSR*s by 15.7 percent (significant).

The high-inflation cases, Brazil and Chile, exhibit a negative relation between monetary and real growth and possible autocorrelation. As shown in Table 1, a pooled regression of the high- and moderate-inflation cases in the form of (3′) using separate country constants (or (3) using in addition separate lag coefficients and time trends) not only yields an insignificant coefficient of unexpected money growth, but an *SSR* more than 21 percent (15 percent) larger than the sum for the individual regressions. The *F*-statistic for the pooling restriction is 3.36 with (5,105) degrees of freedom for (3′) (or 4.14, (5,95) for (3)), which exceeds the critical value and thereby rejects the hypothesis that the reaction to unexpected inflation is the same in all countries.

The significant difference between the two sets of countries is not surprising. An examination of the Brazilian and Chilean money supply processes, preliminary to

---

tance of complete stock adjustment. On the other hand, the annual average figure for money tends to be close to the June figure and the regression results would probably not be greatly worsened by substitution of the second for the first.

TABLE 2—OUTPUT EQUATIONS USING AUTOREGRESSIVE PREDICTOR OF MONEY GROWTH

	Brazil 1952–74	Chile 1952–70	Colombia 1954–74	Mexico 1950–74	Peru 1954–74
$(5')DY_t=a_2'+a_1DMR_t;\ DMR_t=DM_t-\hat{DM}_t=u_{t4}$					
a_2'	.069^{a}	.035^{a}	.051^{a}	.062^{a}	.051^{a}
	(11.1)	(8.3)	(17.9)	(18.9)	(9.5)
a_1	−.045	−.070	.16^{a}	.12	.02
	(.7)	(1.1)	(2.7)	(2.0)	(.2)
DF	21	17	19	23	19
SSR	.0186	.00567	.00318	.00635	.0117
SEE	.0298	.0183	.0129	.0166	.0248
R^2	.02	.07	.24	.15	.00
DW	1.21^{b}	1.38^{b}	1.44	1.93	1.22^{b}
$(4)DM_t=d_0+d_1DM_{t-1}+d_2DM_{t-2}+d_3DM_{t-3}+u_{t4}$					
d_0	.10	.21^{a}	.27^{a}	.14^{a}	.046
	(1.7)	(2.4)	(3.1)	(3.6)	(1.0)
d_1	.55^{a}	.71^{a}	.02	.02	.42
	(2.5)	(2.7)	(.1)	(.1)	(1.7)
d_2	.03	−.36	−.32	−.17	.06
	(.1)	(1.3)	(1.1)	(.8)	(0.2)
d_3	.11	−.01	−.33	^{d}	.25
	(.5)	(.1)	(1.5)		(1.1)
DF	19	15	17	22	17
SSR	.1851	.0853	.0482	.0732	.0855
SEE	.0987	.0754	.0532	.0576	.0710
R^2	.46	.37	.14	.03	.27
DW	1.96	1.87	1.77	2.10	1.95
t of c^{c}	.1	.1	.1	.1	.2

Sources and *Notes*: See Table 1.

[a] Significant at the 95 percent level.

[b] Possible autocorrelation at the .025 level.

[c] t of c refers to the t-statistic of the coefficient of u_{t-1} in the regression $u_t=cu_{t-1}$.

[d] $t<.1$

calculating *DMR*, indicates that they are significantly different than a random walk with trend (see Table 2). Using (3) or (3′) thus imposes an incorrect money supply process and therefore irrational expectation. In the other three cases the random walk hypothesis is acceptable. All five money supply processes are stable and, except for Brazil and Peru, exhibit at least one negative or corrective coefficient.

As shown in Table 2, equation (5′) (or the statistically indistinguishable (5)) yields negative insignificant coefficients of unexpected money growth and possible autocorrelation in the cases of Brazil, Chile, and Peru. The moderately inflationary countries exhibit similar results for (3′) and (5′) (or (3) and (5)) in terms of coefficients, significance levels, and autocorrelation, as might be expected given the reasonableness of the random walk hypothesis for money supply. (Compare Tables 1 and 2.)

It also seems reasonable to explore Sargent and Wallace's hypothesis regarding rational expectations based on past inflation. This is particularly true since in Brazil and Chile lagged inflation is a better predictor of monetary growth than lagged monetary growth.[16] (The ratio of standard errors

---

[16] For Chile and Brazil, throughout this paper, lagged inflation refers to the period December to December. This information is one-half year more up-to-date than that provided by lagged changes in the average money stock (centered at June 30). Lagged inflation, so defined, is not contemporaneous with output, except that the December price level is announced in January or February. If prices adjust rapidly to variations in the rate of monetary expansion, or if many prices in the

TABLE 3—BRAZIL AND CHILE OUTPUT EQUATIONS: INFLATION AS A PREDICTOR OF MONEY AND POOLED EQUATION USING TWO MONEY SUPPLY PROCESSES

	Brazil[a] 1952–74	Chile[a] 1952–70	Five-Country Pooled
	(5′) $DY_t = a_2' + a_1 DMR_t$, $DMR_t = DM_t - \hat{DM}_t$		
a_2'	.069[c]	.035[c]	[b]
	(11.2)	(8.2)	
a_1	.086	.073	.111[c]
	(.9)	(.9)	(3.5)
DF	21	17	109
SSR	.0184	.0058	.0482
SEE	.0269	.0185	.0210
R^2	.03	.04	.29
DW	1.19[d]	1.24[d]	–
	$DM_t = d_0 + d_1 DP_{t-1} + d_2 DP_{t-2} + d_3 DP_{t-3} + u_t$		
d_0	.132[c]	.243[c]	
	(4.2)	(6.3)	
d_1	.66[c]	.50[c]	
	(5.4)	(3.5)	
d_2	.11	–.06	
	(.8)	(.4)	
d_3	–.04	–.16	
	(.3)	(1.2)	
SSR	.0858	.0474	
SEE	.0672	.0562	
R^2	.75	.65	
DW	1.68[e]	1.63[e]	

Sources and *Notes*: See Table 1.

[a] The prediction equation using mid-year prices yields an *SEE* of .0761 for Brazil and .0696 for Chile. The corresponding values of a_1 (and t-statistic) in (5′) are .03 (.3) and –.04 (.6).

[b] The country growth rate constants (and t-statistics) are .069 (15.8), .034 (7.2), .033 (4.9), .036 (5.6) and .049 (8.7). A pooled regression, weighted by the inverse of individual country *SEE*s, increased the total *SSR*s by .42 percent. A regression in the level form yielded $a_1 = .101$ (2.9) and an *SSR* of .0424, relative to a sum of *SSR*s of .0419 for the individual cases; the weighted pooled regression increased *SSR*s by 1.21 percent. See my earlier paper, Table IV, for more details.

[c] Significant at the 95 percent level.

[d] Possible autocorrelation at the .01 level.

[e] Possible autocorrelation at the .025 level.

(*SEE*) shown in the lower half of Table 2 to those of Table 3 is 1.5 and 1.3, respectively.) Moreover, the hypothesis that past money growth "causes" future inflation is rejected using a Sims test while the hypothesis that past inflation affects present money growth

---

index are government controlled and tend to be adjusted in December, then, speaking intuitively, December-to-December inflation should provide better predictions of year-to-year money growth than either lagged money growth or annual inflation rates. This hypothesis seems borne out: predictions of monetary growth with inflation rates calculated from past annual averages of prices are only slightly better in terms of standard errors than those made with past monetary growth rates, and the corresponding output equations are much worse than those in Table 3. December money supply figures are an alternative source of up-to-date information. Unfortunately these are distorted by the large and highly variable year-end bulge in Latin American money supplies which introduces spurious variance into the independent variable and reduces its usefulness as a predictor. Experiments using these data yielded results in equation (5) and in explaining June-to-June money growth which were similar to those obtained in Table 2 with annual averages of monetary growth. For Colombia and Mexico the use of year-end or annual averages of inflation made no difference in the results. The use of year-end money reduced all significance levels in the case of Colombia.

cannot be rejected.[17] For Colombia, Mexico, and Peru, monetary growth is predicted slightly better by lagged monetary growth than by past inflation, but neither regression is really significant. (The corresponding ratio of *SEE*s is .96, .98, and .93, respectively. See Table III in my earlier paper for details.) Sims' tests for these three countries show no clear direction of causation.[18] It is also worth noting that all five countries exhibit at least one negative corrective coefficient of past inflation.

As shown in Table 3, unexpected monetary growth does have a positive (insignificant) effect on real growth in Brazil and Chile, when past inflation is used to predict monetary growth.[19] The use of lagged inflation also makes the predictability of Brazilian and Chilean monetary growth, as measured by the *SEE*, comparable to the standard error of monetary growth for the three moderately inflationary countries (see Table 1). Lucas' model has the strong implication that in such cases the response to unexpected inflation should be about the same. Therefore, the five-country pooled regression again was run, retaining DM_t as the independent variable in the medium-inflation countries, but using the residuals from the lagged inflation predictor of DM_t in the high-inflation countries.

As shown in Table 3, the pooled equation in the form of (5′) with separate country constants (or (5) with separate constants, lag coefficients, and trends) yields a coefficient of .11 (.10) with a *t*-statistic of 3.5 (2.9). The pooled *SSR* is only about 1 percent greater than the sum of the individual country regressions, yielding an *F*-statistic for the pooling restriction of only .13 (5,105) (.27 (5,95)). Thus the hypothesis of similar coefficients across countries is very acceptable, providing strong support for Lucas' theory. Moreover, the individual and pooled country regressions indicate a 95 percent confidence level of about (0.04, 0.16) for the coefficient a_2, which is below the 95 percent confidence interval for all of Lucas' noninflationary countries except Austria, Honduras, and Sweden, again supporting his hypothesis. Finally, the pooled growth form is acceptable since the restrictions of the growth form imply an *F*-statistic of less than 1.4 when compared with either the pooled, level form or the individual country regressions in level form.

III. Empirical Results using Lagged Inflation to Predict Inflation and Explain Output

Table 4 presents estimates of equation (7′) along with statistics on the average rate of inflation, its standard deviation and the *SEE* using DP_{t-1} as a predictor. Again this growth form cannot be distinguished from one in levels, implying in this case that a single change in money has no permanent effect, but the hypothesis cannot be rejected that a once and for all change in monetary growth could raise income above

---

[17] The *F*-statistic for constraining the three coefficients of lagged values of *DM* to zero, relative to the regression

$$DP_t = c_1 + \sum_{t-3}^{t+3} e_i DM_i$$

was only 1.35 with (3,11) degrees of freedom for Brazil and 2.33 (3,7) for Chile, both below the critical values. By comparison the *f*-statistics for constraining the future coefficients to zero are 8.32 (3,11) for Brazil and 5.05 (3,7) for Chile, both far above the critical values. This is confirmed by running regressions of the form

$$DM_t = c_2 + \sum_{t-3}^{t+3} f_i DP_i$$

Constraining the three future coefficients to zero has no effect on the Brazilian regression; the *F*-statistic for Chile is only .74 (3,7). Constraining the lag coefficients to zero yields *F*-statistics of 7.10 (3,11) for Brazil outside the critical region, but only 1.00 (3,7) for Chile. Results using two leads and lags were similar, except that dropping the lagged price coefficients in the regressions of *DM* as a function of DP_t's produced a significant fall in the *SSR* for Chile. Essentially the same results also were obtained using inflation in the annual average of prices. All data were unfiltered.

[18] In no case was the *F*-statistic greater than 1.64, and only four exceeded 1.0. By comparison, the critical value for 3,13 degrees of freedom is 3.49; for 3,9 it is 3.86. Again all data were unfiltered. Inflation refers to changes in the annual average of prices.

[19] The other three cases are not shown, but the output equation results are about the same as (5′) using monetary growth for Colombia and Mexico, and as (3′) for Peru, in terms of coefficients and fits. For details, see my earlier paper, Table III.

TABLE 4—OUTPUT GROWTH AS A FUNCTION OF MONEY GROWTH LESS LAGGED INFLATION

	Brazil[a] 1952–74	Chile[a] 1952–70	Colombia 1951–74	Mexico[b] 1950–74	Peru 1951–74	Pooled
			$(7')DY = a_2^{*\prime} + a_1^*(DM_t - DP_{t-1})$			
a_1^*	$.058^d$	$.030^d$	$.047^d$	$.052^d$	$.046^d$	^{c}
	(7.9)	(7.6)	(13.1)	(4.7)	(6.8)	
$a^{*\prime}_2$	166^d	$.109^d$	.067	$.156^d$	.075	$.12^d$
	(2.2)	(2.9)	(1.9)	(3.8)	(1.2)	(5.0)
DF	21	17	22	23	22	109
SSR	.0154	.0041	.0042	.0046	.0144	.0437
SEE	.0271	.0155	.0140	.0142	.0256	.0200
R^2	.19	.32	.14	.38	.06	.36
DW	1.59	1.71	1.51	2.33	1.66	–
Standard Deviation of Inflation	.146	.140	.081	.051	.034	–
Average Inflation	.259	.281	.109	.059	.085	–
$\sqrt{\frac{(P_t - P_{t-1})^2}{DF+1}}$	.134	.155	.079	.057	.041	–

Sources and *Notes*: See Table 1.

[a] Brazil and Chile use year-end wholesale price data in computing DP_{t-1}. The corresponding regressions using annual averages are

$$DY_t = \underset{(8.4)}{.062} + \underset{(1.7)}{.12}\,(DM_t - DP_{t-1}),\ SSR = .0168,\ R^2 = .12,\ DW = 1.54$$

and

$$DY_t = \underset{(7.4)}{.033} + \underset{(1.4)}{.048}\,(DM_t - DP_{t-1}),\ SSR = .0054,\ R^2 = .10,\ DW = 1.48$$

[b] Mexico uses Consumer Price Index. Results using Wholesale Price Index are similar.

[c] Respective individual country growth rate constants (and t) were .062 (10.7), .030 (8.2), .044 (14.3), .045 (8.4), .055 (17.7). A pooled regression, weighting individual country samples by the inverse of their standard errors, resulted in an SSR which was only 3.05 percent larger than the sum of the individual cases. For details see my earlier paper, Table VII.

[d] Significant at the 95 percent level.

trend.[20] Output equations in the form of (9) or (9′), using more complicated autoregressive predictors of inflation, were inferior to those using either lagged inflation or autoregressive predictors of money and are not presented here.[21]

---

[20] Regressions of the level of output on this measure of unanticipated inflation, a trend, and the lagged level of output, generally yielded the following: coefficients of the third variable which differed from unity by the statistically insignificant amount of 2 percent or less; statistically insignificant trend coefficients; coefficients of unanticipated inflation which were quite similar to those obtained in the growth equation; and Durbin-Watson statistics which would allow rejection of the hypothesis of autocorrelation, although the test is somewhat biased.

[21] All countries displayed one negative coefficient, although only in the case of Chile was it significantly different from zero. All countries displayed at least one significant coefficient except Mexico and Peru, but in all cases except Brazil it was impossible to reject the constraint that the sum of the coefficients of lagged inflation was zero. In Brazil, Mexico, and Peru, an adaptive process could not be rejected, although the power of the test was quite low for the last two cases. The output equation using this process, (9′), was much worse than (7′) for Brazil and Mexico, slightly worse than (7′) for Chile, and about the same as (5′) for Colombia and about the same as (3′), (7′), or (5′) using inflation for Peru. For more details, see my earlier paper, Table VI.

Simple lagged inflation provides a very good reduced-form predictor of inflation. The proportionate effect on output growth of (the reduced form of) unexpected inflation remains about .1, but now is statistically significant in three cases and nearly so for Colombia. The Durbin-Watson statistics exceed 1.4, suggesting that the growth effects of unexpected changes in inflation tend to be solely contemporaneous. In particular, for Brazil and Chile, monetary growth less lagged inflation (December to December) provides a more significant explanation of deviations from trend growth than any of the other more complicated forms. The previous autocorrelation in the residuals is now eliminated. Mexican output growth is also best explained by (7′); for the case of Colombia (5′) and for Peru (3′) are the best output equations.

These are interesting results since the various reduced forms for output differ only in the weights assigned to past money growth and to the "errors" in a growth form of (2). Nonetheless, in rapid inflation the differences in weights implied by the use of lagged inflation yield a much better reduced form proxy for predicted inflation than the roundabout process of proxying unexpected inflation by the errors from predictions of monetary growth. To some degree this result could be attributed to more up-to-date information, but even the use of changes in the average price level in (7′) yields somewhat better results than (3′) or (5′). (See Table 4, fn. a.) At the least, the improvement in the output equation provides additional confirmation of the point that "rational" predictors of inflation involve past inflation if monetary growth policy is made on the basis of past inflation.

In less rapid inflation, such as Colombia and Peru, in which both monetary growth and inflation more closely approximate a random walk with trend, there seems to be little to choose between the different weighting schemes, and even a random walk with trend works. The case of Mexico is something of an anomaly. By far the best proxy of inflation in the output equation is $DM_t - DP_{t-1}$, despite the random walk nature of the monetary process. Possibly this is due to Mexico's greater reliance on a fixed effective exchange rate, which would make money more of an endogenous variable.

The similarity of these five-country coefficients suggest a test of their equality. As shown in Table 4, no statistically significant difference in individual country coefficients is indicated by a comparison of the pooled *SSR* with the *SSR*s from the individual equations; the *F*-statistic is only .54 (5,105 *DF*). A good approximation continues to be that an unexpected ten percentage point increase in the rate of monetary expansion yields about a one percentage point increase in the growth rate. This extra growth was "paid for," within the sample, as money growth returned to its usual rate and lagged inflation overshot the actual rate.

While at first glance these results seem in accordance with the earlier pooled results, they actually raise some questions and suggest the need for further theoretical and empirical work. Under Lucas' hypothesis, the similar reaction to unexpected inflation discussed in Section III made sense; all five countries experienced equally predictable money growth. Statistically similar country coefficients are also obtained for the response to monetary growth less lagged inflation, and the fits are somewhat better than those of Section III. However, as shown in Table 4, lagged inflation is far from an equally good predictor of inflation in all countries, which casts some doubt on the hypothesis that responsiveness to unexpected shocks in aggregate demand is inversely related to the predictability of shocks in aggregate demand. Of course, the Lucas model is not strictly comparable to the one used here and the differences in the quality of the forecast seem to be between Peru and Mexico and the other three countries.

IV. Experiments with Changes in Monetary Regimes—Brazil

Changes in foreign circumstances or local government policy may have produced substantial differences in the autoregressive structures of inflation within the periods studied. If this is true, then lagged inflation may be a better proxy for expected inflation

TABLE 5—BRAZILIAN OUTPUT GROWTH; TWO MONETARY REGIMES[a]

1952–74	$DY_t = \underset{(9.3)}{.060} + \underset{(2.7)}{.090}\, DMR_t$
$SEE = .0250$	$SSR = .0142 \quad R^2 = .26 \; DW = 1.69$
1952–65	$DM_t = \underset{(1.6)}{.05} + \underset{(1.8)}{.36}\, DM_{t-1} + \underset{(3.8)}{1.03}\, DM_{t-2}, \quad DMR_t = DM_t - \hat{DM}_t$
$SEE = .0400$	$SSR = .0176 \quad R^2 = .94 \quad DW = 2.36$
1967–74	$DM_t = \underset{(1.2)}{.19} + \underset{(.5)}{.21}\, DM_{t-1} + \underset{(1.1)}{.19}\, DM_{t-2}, \quad DMR_t = DM_t - \hat{DM}_t$
$SEE = .0510$	$SSR = .0131 \quad R^2 = .19 \quad DW = 1.30^b$
1966	$\hat{DM}_t = .3775, \; (DM_t = .2275), \; DMR_t = .15$
Total *SSR* for expectation equations = .0612	

[a] Equation (5′) using past money growth for prediction.

[b] Indicates possible autocorrelation at .025 level.

than the results of a single prediction equation imposed over the whole period. While this may be remedied by using subperiod prediction equations, the erratic nature of Latin American policymaking reduces the usable sample size and the small number of annual observations prohibits use of ARIMA techniques. One possible exception is the case of Brazil which experienced a military coup in 1964. After 1965 the military government followed essentially similar monetary policies until 1974 (but see Samuel Morley).

Table 5 illustrates the best results from subdividing the sample into pre- and post-1965 autoregressive inflation structures; equations using lagged inflation as a predictor are slightly worse. The *SSR*s for the predictions of money growth are roughly a third of those obtained in Table 2, a statistically significant difference.[22] In the first period, monetary growth and inflation are unstable. The sum of the coefficients of lagged money growth exceed one, a statistically significant difference, even if the constant is omitted. In the second period inflation is stable and the money supply process approximates a random walk with trend.

The corresponding output equation (5′) dominates all previous results with a strongly significant coefficient for unexpected money growth, but one which is not significantly different than those obtained for the other four countries. This not only provides additional cross-sectional support but confirms Lucas' hypothesis in two additional ways. First, the predictability of inflation is about the same in both periods, so the response coefficient should be about the same but lower than in the developed countries. Second, the output equation fits much better when we take into account the different formation of expectations under the two different money supply processes. Unfortunately, this methodology could not be applied to other cases, for example, Chile and Peru, owing to the relatively short life of objectively identifiable monetary regimes.

V. Persistent Supply Shocks versus Lagged Effects of Unexpected Inflation

Recent theoretical work by Lucas (1975) and empirical work by Barro (1977, 1979) suggested that unexpected increases in inflation may have persistent effects. In particular, Barro finds significant coefficients of contemporaneous and two lagged values of unexpected changes in money in the United States, and jointly significant coefficients in the case of Mexico. The coefficients tend to have a triangular pattern. There obviously is some interest in seeing whether these results can be replicated.

Persistent "supply" shocks such as crop failures, natural disasters, world price

---

[22] The *F*-statistic exceeds 5, with (4,17) degrees of freedom; a similar result is obtained for the constraint of two inflationary regimes.

TABLE 6—TESTS FOR LAGGED EFFECTS OF UNEXPECTED INFLATION

$$Y_t = a_0 + a_1 X_t + a_2 \text{ (Time)} + a_3 Y_{t-1} + a_4 X_{t-1} + a_5 X_{t-2} + a_6 X_{t-3}$$

	Brazil^{c} 1952–74 $X_t=(DM_t-DP_{t-1})$	Chile 1954–70 $X_t=DMR_t^{d}$	Colombia 1956–74 $X_t=DMR_t^{e}$	Mexico 1950–74 $X_t=(DM_t-DP_{t-1})$	Peru 1956–74 $X_t=DMR_t^{e}$
a_0	.54	$-.86^{a}$	$-.19^{a}$	$-.88$	$-.37$
	(1.2)	(2.5)	(2.2)	(1.6)	(0.7)
a_1	.135	.108	.098	$.130^{a}$	.068
	(1.6)	(1.2)	(1.4)	(2.9)	(0.6)
a_2	.010	$.015^{a}$	.009	.013	.010
	(1.3)	(2.5)	(1.4)	(1.7)	(0.9)
a_3	.851	$.631^{a}$	$.854^{a}$	$.794^{a}$	$.796^{a}$
	(1.3)	(3.9)	(6.8)	(6.6)	(4.0)
a_4	.114	.021	$-.069$	.043	.046
	(1.3)	(0.2)	(0.9)	(1.0)	(0.4)
a_5	.161	^{j}	.004	$-.001$	.044
	(1.9)		(0.1)	(0.3)	(1.5)
a_6	.004	^{j}	^{j}	^{j}	^{j}
	(1.0)				
DF	16	12	13	19	13
SSR	.01096	.002324	.001657	.00355	.00825
SEE	.0262	.01454	.01129	.0137	.0252
DW	1.71	1.66	2.07	2.29	1.19^{b}
F of H_1^{f}	1.62	.03	1.64	1.93	.13
F of H_2^{g}	43.62^{i}	16.20^{i}	46.54^{i}	43.0^{i}	15.8^{i}
F of H_3^{h}	1.30	3.62^{i}	2.60	1.15	.40

Sources and *Notes*: See Table 1.

[a] Significant at the 95 percent level.

[b] Indicates possible autocorrelation at the .025 level.

[c] Fit using lagged money was slightly better but all coefficients of lagged money were negative. The results of F-tests on constraints yielded similar results.

[d] $DMR_t = DM_t - \hat{DM}_t,\ \hat{DM}_t = \text{constant} + \sum_{i=1}^{i=3} d_i^{*} DP_{t-1}$

[e] $DMR_t = DM_t - \hat{DM}_t,\ \hat{DM}_t = \text{constant} + \sum_{i=1}^{i=3} d_i DM_{t-1}$

[f] $H_1 = a_4$ to $a_6 = 0$

[g] $H_2 = a_3 = 0$

[h] $H_3 = a_2,\ a_4$ to $a_6 = 0,\ a_3 = 1$

[i] Constraint significantly affected SSR as measured by F-statistic at the 95 percent level.

[j] $t < .1$

changes, and trade-exchange regimes also may be important sources of persistent output variations in the resource-based, externally oriented countries considered here. Casual empiricism suggests output is above or below "normal" for fairly long periods. This is confirmed by the large estimated coefficients of lagged output cited above. Moreover, such shocks may have complex interactions with monetary growth and inflation. Thus, neglect of some measure of these shocks would bias the significance of lagged money growth in output regressions.[23] One way to reduce this problem would be to include lagged values of output in regressions of output on lagged demand disturbances, and then test for significance of the two groups of variables.[24] While the inclusion of lagged output naturally would reduce the significance of lagged demand disturbances, the previous results suggest there is enough independent, contemporaneous variation in the series to permit estimation of the significance of both variables.

---

[23] Lagged output might also influence current predictions of inflation and/or monetary growth. However, this effect cannot be separated from a lagged direct effect on output.

Table 6 presents the joint hypothesis regression of output on time, lagged output (more than one lagged value of output was insignificant), and the current and lagged unexpected inflation series which yielded the best fit. Generally the unexpected inflation series was derived from the best predictor used earlier, but the results of the tests are robust to all methods of predicting inflation previously used.

Although the coefficients of contemporaneous and lagged unexpected inflation display a triangular pattern when lagged output is omitted, this disappears when lagged output is included. To test for the independent influence of lagged unexpected inflation and of lagged output, a comparison was then made between the *SSR* obtained, omitting one or the other set of variables and that obtained for the joint hypothesis. Finally, the *F*-statistic for the growth rate form is shown for purposes of comparison.[25]

In all cases, omitting the lagged value of output significantly worsened the fit of the output equation. Thus lagged output enters independently of contemporaneous and lagged demand disturbances. However, omitting the lagged values of unexpected inflation had an insignificant effect. Thus, while in some cases lagged values of unexpected inflation may be jointly significant in explaining output if lagged output is omitted, a large part of their significance is due to their correlation with lagged output. Once lagged output is included, lagged unexpected inflation has an insignificant independent influence. Finally, the (constrained) growth rate form fits the data between two and four times better than regressions including only contemporaneous and lagged values of unexpected inflation and a trend. These results indicate that if demand disturbances have lagged effects, then it is mainly through their initial effect on output. Further, output persistently deviates from trend, independently of the measures of demand disturbances used here, indicating strong supply effects.

VI. Conclusions

This paper has investigated the longstanding question of the relation between inflation and growth in Latin America by applying a rational or consistent expectations approach to five inflationary Latin American countries. As a rule of thumb, ten percentage points of unexpected inflation raise output about one percentage point above trend, or yield one percentage point of extra growth, when the data are interpreted in a growth rate framework. The growth formulation cannot be rejected with the present set of data.

The use of different simple hypotheses regarding the money supply process can equalize the predictability of inflation in the five countries under consideration. This predictability is low by Organization for Economic Cooperation and Development (OECD) standards. The corresponding estimated coefficient of unexpected inflation is statistically indistinguishable across the five countries and significantly lower than in OECD countries, supporting Lucas' hypothesis that reaction to unexpected inflation is inversely related to its predictability.

Many different processes for forming expectations of inflation work approximately as well in explaining output in medium inflation countries. In Brazil and Chile, money growth is predicted much better by past inflation than past monetary growth. The former predictions also work better in the output equations. Also, at least one lagged value of inflation or monetary growth seems to have a negative correlation with current inflation, casting some doubt on adaptive expectations. Experiments with Brazilian data indicate fairly rapid adjustment to new monetary supply processes, as measured by predictability of inflation and fit of the output equations.

---

[24] This test was suggested by Robert King.

[25] Differences in sample size mean the constrained *SSR* is not always shown in earlier tables.

Judged solely by the significance of the output equation, lagged inflation is the best predictor of inflation in three of the five cases, and nearly as good in the other two. The pooled regression supports the aforementioned 10 percent rule of thumb, but does not permit rejection of the hypothesis that the countries react similarly to unexpected inflation. However, this result casts some doubt on Lucas' hypothesis, for lagged inflation is far from being an equally good predictor of inflation in the five countries.

Finally, most of the output effect of unexpected monetary disturbances seems to be contemporaneous. While there are persistent deviations in output from normal levels, as the growth form would imply, these seem to be related more to supply shocks than to the lagged effects of unexpected changes in money. Future work will explore how these shocks might be related to the foreign sector.

REFERENCES

M. Arak, "Some International Evidence on Output-Inflation Tradeoffs: Comment," *Amer. Econ. Rev.*, Sept. 1977, *67*, 728–30.

R. Barro "Unanticipated Money Growth and Employment in the United States," *Amer. Econ. Rev.*, Mar. 1977, *67*, 101–15.

______, "Money and Output in Mexico, Colombia, and Brazil," in Jere Behrman and James Hanson, eds., *Short-Term Macroeconomic Policy in Latin America*, Cambridge, Mass. 1979.

______, "The Short-Run Relation Between Output and Inflation: Is There a Phillips Curve in Colombia?," Departamento Nacional de Planeación, Bogotá 1974.

______ **and S. Fischer,** "Recent Developments in Monetary Theory," *J. Monet. Econ.*, Apr. 1976, *2*, 133–67.

W. Baer and P. Beckerman, "Indexing in Brazil," *World Develop.*, Oct./Dec. 1974.

P. Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *The Quantity Theory of Money*, Chicago 1956.

R. Campos, "Two Views on Inflation in Latin America," in Albert Hirschman, ed., *Latin American Issues*, New York 1961.

Carlos Diaz Alejandro, *Essays on the Economic History of the Argentine Republic*, New Haven 1970.

______, *Foreign Trade Regimes and Economic Development: Colombia*, New York 1972.

A. Diz, "Money and Prices in Argentina," in David Meiselman, ed., *Varieties of Monetary Experience*, Chicago 1970, 111–23.

D. Dutton, "A Model of Self-Generating Inflation," *J. Money, Credit, Banking*, Nov. 1971, *3*, 245–62.

R. Fernandez, "An Empirical Inquiry on the Short-Run Dynamics of Output and Prices," *Amer. Econ. Rev.*, Sept. 1977, *67*, 595–609.

______ **and J. Hanson,** "Las Interrelaciones Del Corto Plazo entre Inflación, Producción y Empleo en América Latina," in *Planificación de Corto Plazo*; *La Dinámica de los Precios, el Empleo y el Producto*, ILPES Cuaderno no. 25, Santiago 1977.

Ricardo Ffrench-Davis, *Politicas Económicas en Chile, 1952–1970*, Santiago 1973.

J. Hanson, "The Short-Run Relation Between Growth and Inflation in Latin America: A Quasi Rational or Consistent Expectations Approach," *Int. Finance Disc. Papers*, No. 118, May 1978.

A. Harberger, "The Dynamics of Inflation in Chile," in Carl Christ, ed., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Honor of Yehuda Grunfeld*, Stanford 1963, 219–50.

______, "El Rol de los Factores Fiscales en la Inflación Uruguaya," *Cuadernos de Economia de la Universidad Católica de Chile*, Año 12, No. 37, dic. 1975, 33–46.

R. King, "A Note on Expectations and Tests of Temporal Causality," unpublished paper, Brown University 1977.

R. Lucas, "Econometric Testing of the Natural Rate Hypothesis," in Otto Eckstein, ed., *Conference on the Econometrics of Price Determination*, Washington 1970, 50–59.

______, "Some International Evidence on the Output-Inflation Tradeoffs," *Amer. Econ. Rev.*, June 1973, *63*, 326–34.

______, "An Equilibrium Model of the Business Cycle," *J. Polit. Econ.*, Dec. 1975, *83*, 1113–44.

——— **and L. A. Rapping,** "Price Expectations and the Phillips Curve," *Amer. Econ. Rev.,* June 1969, *59,* 342–50.

S. Morley, "Inflation and Stagnation in Brazil," *Econ. Develop. Cultural Change,* Jan. 1971, *19,* 184–203.

J. Muth, "Optimal Properties of Exponentially Weighted Forecasts," *J. Amer. Statist. Assn.,* June 1960, *55,* 299–306.

E. Phelps, "Money Wage Dynamics and Labor Market Equilibrium," *J. Polit. Econ.,* Aug. 1968, Part II, *76,* 678–711.

T. Sargent and N. Wallace, "Rational Expectations and the Dynamics of Hyperinflation," *Int. Econ. Rev.,* June 1973, *14,* 328–50.

E. Sheehy, "The Dynamics of Inflation in Latin America: Comment," *Amer. Econ. Rev.,* Sept. 1976, *66,* 692–94.

Enrique Sierra, *Tres Ensayos de Estabilización en Chile,* Santiago 1970.

C. Sims, "Money, Income, and Causality," *Amer. Econ. Rev.,* Sept. 1972, *62,* 540–53.

R. Vogel, "The Dynamics of Inflation in Latin America, 1950–1969," *Amer. Econ. Rev.,* Mar. 1974, *64,* 102–14.

Susan Wachter, *Latin American Inflation,* Lexington 1976.

International Monetary Fund, *Int. Financ. Statist.,* various years.

United Nations, *Yearbook of National Income Accounts Ststistics,* New York, various years.

Interest Rates, Inflation Expectations, and Spurious Elements in Measured Real Income and Saving

By GREGORY V. JUMP*

The high rates of price inflation which began early in the last decade and continue today have been accompanied by some unusual movements in aggregate economic variables. Rates of personal saving rose to record high levels in the mid-1970's throughout the industrialized world. With the notable exception of the United States, these high saving rates have persisted into this decade. Profit rates fell below the averages of earlier decades with the advent of high inflation, though there remains some doubt as to whether this represents something other than a cyclical phenomenon.[1] Government deficits increased dramatically in the 1970's. The fact that these developments occurred during a period of rising inflation is not purely coincidental. The purpose of this paper is to point out a general set of dependencies which directly relate the foregoing events to increases in the expected rate of price inflation. The dependencies exist because of peculiarities in conventional measures of income.

Conventional accounting practice is to include interest receipts, net of interest expenses, as part of measured income and to exclude capital gains and losses. When the expected rate of price inflation increases, nominal rates of interest rise by an inflation premium which compensates the owner of a fixed-income asset (at least partly) for the capital depreciation suffered on his asset holdings. This causes measured income to change because the actual capital loss (creditor) or gain (debtor) on the fixed-income asset is not deducted from interest earnings or expenses in computing income. The measured income of creditors will rise with expected inflation while that of debtors will decline. The effects appear in both nominal and real measured incomes. Measured real saving will move in the same direction as measured real income, and saving rates will be similarly affected.

If price inflation was completely neutral in its impacts on the economy and could be perfectly anticipated, the described effects would still occur. Any movements in measured real income and saving as a result of changes in inflation under these circumstances would have to be regarded as spurious; they would not reflect any changes in economic behavior or well-being.

In this paper it will be argued that conventional measures of income and saving contain spurious elements related to inflation expectations even in a world in which inflation is nonneutral in its impacts and cannot be perfectly anticipated. It will also be argued that a number of recent observations which have puzzled economists, such as those mentioned in the introductory paragraph, are partly attributable to spurious movements associated with the rapid rise in inflation expectations during the early 1970's.

Some connections between inflation and income measurement problems have been identified in isolated instances in the recent literature. A number of writers (for example, see James Pesando and Samuel Rea, Jr.) have observed that in order to maintain the real value of their assets constant during an inflationary period, consumers must save an increased fraction of their incomes. This comes close to the recognition of an inflation-related spurious element in measured personal saving. In a different context, Feldstein and Summers (1977) observed that the rate of profit, when measured as the ratio of profits-before-interest payments to capital stock, does not have the downward shift during the early and mid-1970's that appears in the conventional profits-to-capital stock ratio. This too comes close to the identification of an inflation-related spurious element, though these authors did not recognize it as such.[2]

---

*Department of political economy and Institute for Policy Analysis, University of Toronto. Financial support of the Economic Council of Canada is gratefully acknowledged. I would like to thank Jack L. Carr, James E. Pesando, and Samuel A. Rea, Jr., for helpful comments on an earlier draft.

[1] For a discussion of this issue, see the papers by William Nordhaus and by Martin Feldstein and Lawrence Summers (1977).

The remainder of this paper analyzes the nature and magnitude of these and other spurious elements in measured variables. Section I considers the impact of a fully anticipated price inflation on the measured income and saving of an individual in an inflation-neutral setting. The individual's net interest income is also computed using an inflation invariant concept of income proposed by John Hicks. The magnitude of the spurious element in the conventional measure of income is shown to be equal to the algebraic difference between the two income concepts. This provides a measure of the spurious element which is shown to be robust to conditions of uncertainty and nonneutrality.

In Section II the analysis is extended to an economy-wide basis, and measures of inflation-related spurious elements in aggregate real incomes and saving are derived. A general method for computing "inflation-adjusted" income and saving is proposed. The method is applied to one measure of the *U.S.* personal saving rate in Section III.

I. Spurious Inflation Impacts on an Individual's Real Income and Saving

Consider the effects of perfectly anticipated price inflation on an individual whose entire net worth at the start of some time period t = 1 is assumed to consist of a stock of nominal fixed-income assets A_0, carried forward from the end of the preceding time period. Assume that the nominal interest rate earned on these assets during any time period t is i_{t-1}, determined at the end of period t − 1, and that this rate is renegotiated at the end of every period with perfect foresight regarding all future price inflation. Let us also assume that interest receipts are subject to a proportional tax at the rate τ.

Suppose this individual expects to live until the end of time period $t=N$. Any assets, $A_N \geqslant 0$, which he holds at the time of his death will be passed on to his heirs in the form of a bequest. The real value of any such bequest will be (A_N/P_N) at the time of his death.

The individual may be presumed to have a utility function of the form $U = U(C_1, C_2, \ldots, C_N;\ (A_N/P_N))$, where C_t denotes his real consumption during period t.[3] At the start of period 1 he will choose the values of $C_t (t=1,2,\ldots,N)$ and (A_N/P_N) which maximize utility, subject to the restriction that $A_N \geqslant 0$ and the following budget constraint:

$$(1)\quad A_t = (1+(1-\tau)i_{t-1})A_{t-1} - P_t C_t$$

$$t=1,2,\ldots,N$$

The budget constraint is expressed in terms of nominal assets and the nominal rate of interest, but since the arguments in the utility function are all real variables, it would be more appropriate to express it in real terms. This is accomplished by simply dividing both sides of the equation by the price level P_t:

$$(2)\quad \frac{A_t}{P_t} = \left[\frac{(1+(1-\tau)i_{t-1})}{(P_t/P_{t-1})}\right]\frac{A_{t-1}}{P_{t-1}} - C_t$$

$$t=1,2,\ldots,N$$

The term in square brackets in (2) defines the after-tax real rate of interest faced by the individual at the end of period t − 1; or, what is the same thing, it measures the rate of transformation of real saving in period t−1 into real consumption in period t.[4] Letting r_{t-1} denote this after-tax real rate of interest, the budget constraint may be written as

---

[2] They attribute the result to a change in debt-equity ratios, which may be partly correct.

[3] The bequest enters the utility function in real, as opposed to nominal, terms because the individual is presumed to be concerned with the well-being of his heirs. His heirs will be better off the larger the stream of real consumption they can obtain over their lifetimes. This is affected by the real value of the bequest.

$$(3) \quad \frac{A_t}{P_t} = (1+r_{t-1})\frac{A_{t-1}}{P_{t-1}} - C_t;\ t=1,2,\ldots,N$$

Suppose this individual lives in a completely inflation-neutral world where changes in the perfectly anticipated rate of price inflation have no impacts upon real decisions or constraints faced by him or any other economic agent. Suppose, initially, that the rate of price inflation expected to prevail into the indefinite future is zero, so that the future price level will remain at the constant value P_0 established during period $t=0$. For simplicity, assume that the nominal interest rate is also expected to remain constant at the value $i(0)$. The after-tax real rate of interest will also be constant at the value $r=(1-\tau)i(0)$.

Let C_t^* and $A(0)_N$ denote the values of real consumption and nominal bequest, respectively, which maximize the individual's utility under these circumstances.[5] With C_t^* and $A(0)_N$ determined, the budget constraint implies that there is one time profile of real asset accumulation which is consistent with the optimal consumption-bequest plan. Let $A(0)_t$ represent the optimal asset holding at the end of period t which is consistent with this plan.

Under the conditions specified here, the individual will have measured real income, after tax, of $Y(0)_t$ equal to $(1-\tau)i(0)(A(0)_{t-1}/P_0)$ during period t. He will have a personal saving rate of $s(0)_t$, equal to $1-C_t^*/Y(0)_t$ in the same period. Consider how these measured values would differ if, instead of zero price inflation, some other rate of inflation, $p\neq 0$, were to be perfectly anticipated for time period $t=1$ and beyond.

Since the individual is assumed to live in an inflation-neutral world, we know that a change in the value of p can have no effects on the real constraints he faces or the real choices he makes. Indeed, this is the meaning of "neutrality." To translate that into the concepts developed above, a change in the value of p must leave his real budget constraint (equation (3)) unaltered. This can only happen if the after-tax real rate of interest is invariant to such a change. With r unaffected by a change in p, the individual's optimal real consumption-bequest plan will not vary; C_t^* will remain the optimal value of real consumption during period t. Similarly, the *real* value of the bequest in period $t=N$ will remain equal to $(A(0)_N/P_0)$. Of course, the *nominal* value of the optimal bequest will change with p because changes in the rate of price inflation will alter the price level, P_N.

Denote the nominal value of the optimal bequest when inflation occurs at the rate p as $A(p)_N$. From the preceding it is clear that $A(p)_N$ will be equal to $(1+p)^N A(0)_N$. The change in the nominal value of optimal asset holdings will not be restricted to the terminal period. The nominal stock of assets at the end of every time period $t>0$ will change in such a way as to keep the real value of assets at the levels which satisfy the optimal real consumption-bequest plan. Formally,

$$(4) \quad A(p)_t = P_t\left(\frac{A(0)_t}{P_0}\right) = (1+p)^t A(0)_t$$

$$t=1,2,\ldots,N$$

---

[4] The concept of a real rate of interest is, of course, attributable to the classic work of Irving Fisher (1896, 1930). Fisher dealt with a taxless world in which the real rate r is related to the nominal rate i and the rate of price inflation p through the famous Fisher equation: $r=(i-p)/(1+p)$. Fisher emphasized that $(1+r)$ measures the rate at which consumers can transform current real saving into future real consumption. Indeed, that is the basis of Fisher's equation. In subsequent literature, the p in the denominator of the original equation has often been ignored, and the approximation $r \doteq i-p$ is the more familiar equation. In a world where nominal interest receipts are taxed at the rate τ, the rate of transformation of current real saving into future real consumption is still $(1+r)$, but r is in that case the *after-tax* real rate of interest. The analogue to Fisher's equation in a world with taxes is $r=[(1-\tau)i-p]/(1+p)$. This is the basis of the concept of the after-tax real rate used in this paper. The p term in the denominator of this expression is retained throughout the analysis.

[5] The specific nature of the bequest motive has no bearing on the analysis to follow; in particular it does not matter whether the individual chooses to leave a nonzero bequest. What does matter is that he is able to define an optimal real consumption stream which dominates any other feasible stream.

where $A(p)_t$ denotes the new optimal asset holding in nominal terms.

With inflation neutrality, the nominal rate of interest must contain an "inflation premium" which compensates the holder of fixed income assets for both price changes and any changes in tax liabilities which might accompany them. The nominal rate associated with sustained inflation at rate p will be denoted as $i(p)$. Its relationship to r, p, and τ is given by

$$(5)\qquad i(p)=\frac{r(1+p)+p}{(1-\tau)}=i(0)(1+p)+\frac{p}{1-\tau}$$

A change in sustained inflation from zero to p will cause the nominal rate to change by $i(0)p+p/(1-\tau)$. The first component $i(0)p$ compensates for the effects of inflation on the purchasing power of interest receipts. The second, $p/(1-\tau)$, compensates for the capital depreciation incurred in the real value of the stock of fixed assets. This latter term is "marked up" by the tax factor $1/(1-\tau)$ because interest earnings are subject to tax, whereas capital depreciation is not.[6]

When $p\neq 0$, the measured value of the individual's real income in period t is given by

$$(6)\qquad Y(p)_t=(1-\tau)i(p)\frac{A(p)_{t-1}}{P_t}$$
$$=Y(0)_t+p\left(\frac{A(p)_{t-1}}{P_t}\right)$$

There has been a change in the measured value of his real income which must be regarded as spurious, because he is no better or worse off when $p\neq 0$. The magnitude of the spurious change is

$$(7)\qquad d_t=p\left(\frac{A(p)_{t-1}}{P_t}\right)$$

It is clear that $d_t \gtreqless 0$ as $p \gtreqless 0$; a positive rate of inflation will cause a spurious increase in measured real income, while a deflation will cause a spurious decrease.[7]

Note that since inflation is assumed to be perfectly anticipated, the right-hand side of equation (7) may be written as $(p/(1+p))(A(p)_{t-1}/P_{t-1})$. From this it is clear that the magnitude of the spurious element in any time period is independent of the date at which inflation at the rate p first begins. The only time-varying component of d_t is the quantity $(A(p)_{t-1}/P_{t-1})$, the real value of optimal assets at the end of period $t-1$. We have seen that this is determined solely by the optimal consumption-bequest plan and is independent of current and past rates of inflation. A return to anticipated inflation of $p=0$ at the end of period $t-1$ would eliminate the spurious element in period t and beyond.

The individual's measured saving rate in period t will also be affected by the spurious element d_t.

$$(8)\qquad s(p)_t=1-\frac{C_t^*}{Y(0)_t+d_t}$$

When $p>0$, $s(p)_t$ will exceed $s(0)_t$ and the opposite occurs when $p<0$. The individual's measured saving rate will rise and fall with the perfectly anticipated rate of inflation, even though his real consumption is unaffected by such variations.

The spurious element d_t arises because the conventional income measure includes the flow of interest earnings associated with fixed income assets, but it excludes capital gains or losses in the real value of such assets. When p rises, i rises to compensate for capital depreciation, and the flow of interest earnings rises by proportionately more than the price level. Since conventional accounting procedures do not deduct capital depreciation from interest flows, nominal measured income also rises disproportionately to the price level. The size of the capital depreciation in time period t is $pA(p)_{t-1}$ in nominal terms. In real terms, the value is $p(A(p)_{t-1}/P_t)$, precisely equal to the value of d_t identified in (7).

---

[6] The point that the inflation premium embodied in the nominal rate of interest should contain the tax-adjusted term $p/(1-\tau)$ in an inflation-neutral world appears to have first been emphasized by Michael Darby.

[7] The restriction that the results apply only during $t<N$ has been discarded, beginning with equation (6). The spurious element d_t will persist for as long as inflation remains at p. The heirs of the individual will also live in this inflation-neutral world. Their measured real incomes will also contain spurious elements analogous to d_t.

Conventional saving is measured as the difference between income and expenditure. Any spurious movement in measured income must also appear in measured saving. The conventional definition assures that measured saving will be equal to net asset accumulation. When inflation is $p>0$, the individual must accumulate $p(A(p)_{t-1}/P_t)$ more real assets during period t than he would have with $p=0$ in order to offset capital depreciation and achieve the desired real stock $A(p)_t/P_t$ by the end of the period. He therefore "saves" more during inflation in a conventional accounting sense.

The message conveyed by these results is that movements in measured real income may not provide reliable indications of changes in an individual's economic well-being during periods of inflation variability. Movements in measured saving rates may not reflect behavioral changes in the consumption-saving decision.

A. *Inflation-Adjusted Income*

Real world individuals have interest-bearing liabilities as well as assets. They also earn incomes from human capital and other real assets, but inflation appears to give rise to spurious movements only in interest flows. (This issue is examined more closely in the following section.) The preceding analysis suggests an adjustment which could be made to observed values of an individual's total after-tax income to eliminate these spurious affects. Let yd_t denote an individual's conventionally measured real disposable income, let w_t denote his after-tax real income from noninterest sources, and let NA_{t-1} denote his *net* holdings of nominal fixed-income assets ($A(p)_{t-1}$, less all debt liabilities for which interest expenses are deducted in computing conventional income).[8] The term yd_t is defined as the sum of w_t and $(1-\tau)i(p)(NA_{t-1}/P_t)$. The inflation-adjusted measure of income suggested by the earlier results is

$$yd_t^* = yd_t - d_t' \tag{9}$$

where d_t' is defined as $p(NA_{t-1}/P_t)$ in a straightforward extension of equation (7).

The construct yd_t^* would provide a measure of real income which is invariant to changes in perfectly anticipated inflation in an inflation-neutral world. Whether it would be appropriate in a nonneutral world with inflation uncertainty remains to be determined. As a prelude to that question, note that the definition of the after-tax real rate of interest allows yd_t^* to be written as

$$yd_t^* = w_t + (1+p)r(NA_{t-1}/P_t) \tag{10}$$

The adjustment procedure amounts to replacement of the after-tax interest flow which appears in conventional income as $(1-\tau)i(p)(NA_{t-1}/P_t)$ with the alternative measure $(1+p)r(NA_{t-1}/P_t)$. In other words, the adjustment procedure implies that interest flows are better measured using real, as opposed to nominal, rates of interest. With inflation neutrality, the after-tax real tax rate is invariant to p, and equal to $(1-\tau)i(0)$; hence measured income yd_t will be equal to yd_t^* only when the rate of inflation is equal to zero.

It is interesting that Hicks considered some of these same issues more than four decades ago. Hicks did not differentiate between real and nominal rates of interest but was concerned with defining a concept of income which would not vary with changes in price levels. He concluded that "income...must be defined as the maximum amount of money which an individual can spend this week and still expect to spend the same amount in real terms in each ensuing week" (p. 174). An alternative, but equivalent, way of stating this, when the after-tax real rate of interest is constant over time, is to define real income as the maximum amount the individual can consume in any period, and still leave his end-of-period real net worth unchanged.[9]

---

[8] In general, any interest expenses incurred in the process of earning receipts are deducted in computing measured income.

Hicks did not confine his definition to income from fixed-income assets and was convinced that it could never be computed in the real world. Computation would require knowledge of all future receipts and price levels—obviously an impossibility. However, when applied only to interest flows, it is not impossible to compute, even in an uncertain world.[10] In an inflation-neutral world with perfect certainty, it is obviously simple to compute Hicks income from net fixed-income assets NA_{t-1}. The value is $(1+p)r(NA_{t-1}/P_t)$, precisely the concept I have argued is appropriate to inflation-adjusted income, yd_t^*.

We are now in a position to interpret the inflation-adjustment procedure as replacement of conventional net interest income with Hicks income derived from interest-bearing net assets. The spurious element appearing in the conventional measure is the difference between net interest receipts and Hicks income.

Let us extend these ideas to a world of uncertainty and nonneutrality. I shall define as spurious any inflation-related movement in conventionally measured interest income which does not also appear in Hicks income. If a change in inflation expectations causes conventional interest income to change by some amount X_1, and Hicks income by some other amount X_2, then the difference, X_1-X_2, will be regarded as spurious.

B. *Uncertainty and Nonneutrality*

Inflation expectations are seldom held with certainty in the real world. Differences between *ex ante* expectations and *ex post* realizations can and often do occur. Real world inflation is also not neutral in its impact upon real opportunities and decisions. The effects of inflation on real money balances and upon the tax liabilities of corporations provide two examples of how a change in the rate of inflation will cause a redistribution of real resources.[11] In addition, empirical evidence suggests that a change in inflation expectations alters nominal interest rates on approximately a one-to-one basis; that is, a rise in the expected rate of inflation by some amount Δp^e will cause nominal interest rates to rise by approximately this same amount.[12] Under these circumstances, the after-tax real rate of interest facing any investor with a positive tax rate will vary inversely with the expected rate of inflation.

Under conditions of uncertainty, Hicks income during period t is defined as the maximum amount an individual can consume during the period, given the realization P_t, and still expect to be able to consume the same amount in real terms in each ensuing time period.[13] Under the assumption that the individual expects to be able to earn the same *ex ante* after-tax real rate of interest r_{t-1} in period t and all future time periods, Hicks income for an individual with initial net assets of NA_{t-1} is $(1+p_t^e)r_{t-1}(NA_{t-1}/P_t)$.[14] The variable p_t^e is the rate of price inflation expected to prevail during period t, with the expectation held as of the start of the period. The spurious element in conventional income in this case is $p_t^e(NA_{t-1}/P_t)$, analogous to the earlier expression, except that p_t^e is now the appropriate inflation variable. The interesting result is that the spurious element, and hence the appropriate adjustment factor to use in "correcting" measured income, depends only upon the current value of p_t^e. It is independent of the fact that inflation may be nonneutral in its effects upon interest rates.

---

[9] The second statement is essentially the wording used by Milton Friedman in defining his concept of permanent income.

[10] This issue is considered later in this section.

[11] A number of recent studies have investigated the impacts of inflation on the taxation of corporations. The consensus finding is that corporations face higher real tax burdens when inflation rises because depreciation expenses are allowable only at historical costs. For a discussion of this and other issues, see John Shoven and Jeremy Bulow, Glenn Jenkins, and Feldstein and Summers (1978).

[12] See, for example, the studies by William Yohe and Dennis Karnosky, Otto Eckstein and Feldstein, and Feldstein and Summers (1978) using *U.S.* data, and the studies by Carr and Smith and Carr, Pesando and Smith using Canadian data. A paper by Thomas Sargent presents a summary of recent empirical literature.

[13] Hicks was careful to distinguish between *ex ante* and *ex post* versions of his income concept under conditions of uncertainty. The definition in the text is the *ex post* version, based on the realization of the actual price level P_t. This seems the appropriate concept for the issues being examined in this paper, insofar as other components of income can be measured only *ex post*; i.e., after they are realized. The *ex ante* version defines income as the maximum amount the individual *expects* to be able to consume at the start of period t, and still be able to expect to consume that same amount in real terms in each ensuing time period. The *ex ante* version is based on the expectation that P_t will equal $(1+p_t^e)P_{t-1}$.

Does this result make sense? A simple illustration should convince us that it does. Suppose that inflation was formerly zero but rises to some expected rate $p_t^e>0$ at the start of period t. Consider the kind of inflation premium which appears to operate in the real world: the nominal rate of interest rises from $i(0)$ to $i(p)=i(0)+p_t^e$, a one-to-one response. If inflation expectations are realized in period t, the real value of conventionally measured interest income will change in proportion to $(1+p^e-i(0)p^e)/i(0)(1+p^e)$. For values of $i(0)<1$ this will be positive in sign, and conventionally measured real income will rise. But if the individual faces a tax rate $\tau>0$, r_{t-1} will decline as a result of this and he will be worse off. Hicks income will decline. The conventional measure does show a spurious increase under these circumstances, and if we believe that Hicks income provides a better measure of interest income, then the magnitude of the spurious effect is $p^e(NA_{t-1}/P_t)$ as derived above.

How does unexpected inflation affect this result? Suppose $p_t^e=0$, but inflation at the rate $p_t>0$ occurs unexpectedly during period t. The individual is surely worse off. Hicks income will decline by the amount $p_t r_{t-1}(NA_{t-1}/P_{t-1})$. Conventionally measured income will also decline by the amount $p_t(1-\tau)i(0)(NA_{t-1}/P_{t-1})$. Since r_{t-1} will be equal to $(1-\tau)i(0)$ when p_t^e is equal to zero, both income measures will decline by the same amount. There is nothing spurious in this case because inflation was not anticipated. This is also a plausible result.

II. Spurious Elements on an Economy-Wide Basis

Since the spurious elements in the measured real income of an individual are robust to considerations of uncertainty and nonneutrality, we can probe for their existence in an economy-wide setting by considering the effects of perfectly anticipated inflation on a hypothetical, inflation-neutral economy. Consider a closed economy consisting of a government sector, a corporate sector, and a personal sector in which all production is undertaken by corporations. Suppose that all corporate and government debt is held by persons who have no direct debt obligations of their own.

To ensure inflation neutrality, taxes in this economy will be assumed to be levied at the same proportional rate τ on all corporate and personal incomes. The tax system will be assumed to be fully integrated so that double taxation does not occur. Specifically, assume the following:

1) Interest payments on corporate debt are an allowable deduction from corporate

---

[14] There is practically no empirical evidence to judge whether it is reasonable to assume individuals expect to earn the same after-tax real rate of interest in all future periods. The issue is really one involving the term structure of after-tax real rates of interest. Most empirical studies of the term structure have looked only at nominal rates and have found support for the expectations hypothesis. However, there is nothing inconsistent with these findings and the assumption. If the marginal investor does expect to earn the same after-tax real rate over all holding periods, then nominal rates on long-term securities can still differ from those on short-term securities, provided that long-term inflation expectations differ from short-term inflation expectations. In fact, the relationship between long-term rates and current and future expected short-term rates implied by the expectations hypothesis will still hold for nominal interest rates under these circumstances. If the assumption is not valid, Hicks income becomes

$$(1+p_t^e)\bar{r}\left(\frac{1+r_{t-1}}{1+\bar{r}}\right)(A(p)_{t-1}/P_t)$$

where $\bar{r}$ is the after-tax real rate of interest the individual expects to earn in all subsequent periods starting with t+1. Measured income still contains a spurious element in this case, but its identification is more complicated than the expression given in the text.

taxes, but interest receipts of persons are taxed at full rates.

2) Corporate taxes are levied on nominal profits before dividends, and dividend receipts of persons are tax exempt.

3) Corporations are permitted to depreciate investment expenses at replacement costs in computing taxable income.

4) Inflation-related capital gains accruing to equity-owners are not taxed.

To further ensure that inflation is neutral, the after-tax real rate of interest is assumed to be constant at the value r. The nominal rate of interest is, therefore, given by

$$i=\frac{r(1+p)}{(1-\tau)}+\frac{p}{(1-\tau)} \tag{11}$$

where p denotes the perfectly anticipated rate of inflation. To avoid the nonneutrality problems which can arise in the real world where inflation affects the real return derived from holding money balances, we shall assume that money, if it exists at all, yields interest at the nominal rate i. Money is, therefore, indistinguishable from government debt in this economy.

In order to simplify the analysis, let me make the following additional assumptions, none of which is critical to the results:

5) Corporations distribute all receipts in excess of expenses as dividends.

6) The population of this economy is static, real economic growth is zero, and all investment outlays represent replacement expenditures.

The economy will be assumed to have been in a steady-state equilibrium with zero price inflation prior to time period $t=1$. Let $P_0=1.0$ be the price level during this historical interval. A fully anticipated inflation at the rate p begins in period $t=1$ and all wages and prices rise at this rate thereafter. Debt contracts are renegotiated at the end of every time period so that interest payments made in period t are based upon i_{t-1}. As a consequence of these assumptions, the nominal interest rate rises from $i(0)(=r/(1-\tau))$ to the new value i, given in (11) at the end of period $t=0$. It remains at the new value i thereafter.

The onset of inflation will alter the measured values of real income and saving for each of the three sectors of this economy. The government and corporate sectors, which are net debtors, will experience declines in measured income and saving. The personal sector—a net creditor—will experience the opposite effects. Since this economy is inflation neutral, the advent of inflation cannot alter the claims on real resources accruing to any of its participants. Inflation-induced changes in measured income and saving must be spurious.

To identify the spurious elements, let us compare conventionally measured values of income and saving variables before and after inflation with the aid of the following notation. For any nominal income or saving flow X, the symbol X_t will denote its value during period $t \geq 1$, while $\bar{X}$ will be used to denote its value during the pre-inflation interval ($t \leq 0$). Postinflation real values are given by X_t/P_t. Note that since P_0 is assumed to equal 1.0, $\bar{X}$ represents both the nominal and real value of the variable X prior to the onset of inflation. The variable $\bar{X}$ also measures pre-inflation real income using Hicks' definition. Under inflation neutrality, Hicks income cannot change after inflation begins. Any differences between X_t/P_t and $\bar{X}$ which show up in the postinflation period must be spurious, and the magnitude of these spurious effects is measured by the difference $(X_t/P_t-\bar{X})$.

A. *The Government Sector*

It will be assumed, consistent with earlier assumptions, that the government sector was in a zero net balance position prior to the advent of inflation. During this period outstanding government debt was constant at the nominal equals real value $\overline{DG}$, and tax receipts matched the sum of interest payments $i(0)\overline{DG}$, and government purchases of goods and services $\bar{G}$. The nature of the assumed tax system is such that tax revenues are effectively generated by applying the tax rate τ to the sum of nominal gross national product Y, and interest on government debt less nominal investment outlays I. The values of SG, the government budget surplus (+) or deficit (−), before and after inflation are therefore given by the following:

(12a)

$$\overline{SG}=\tau\big(\bar{Y}-\bar{I}+i(0)\overline{DG}\big)-\bar{G}-i(0)\overline{DG}=0$$

(12b)

$$SG_t=\tau(Y_t-I_t+iDG_{t-1})-G_t-iDG_{t-1}$$

The neutrality of this economy guarantees that real values for consumption, investment and government expenditures will be unaffected by the onset of inflation. It follows that the nominal variables I_t, G_t, and Y_t will all grow at the rate of price inflation. No spurious elements will arise in measured values of total *GNP* or any of the components of aggregate demand.[15]

The effect of inflation is to increase the nominal rate of interest by proportionately more than p, so that interest payments on the government debt rise by more than the percentage increase in the price level. Assuming $\overline{DG}>0$, initially, the government is forced into a deficit position with the onset of inflation; i.e., $SG_t<0$. The government must increase its nominal borrowing period by period in order to maintain a constant value of real expenditures and meet its debt obligations. It is easy to verify that the government debt must rise over time in accordance with

$$DG_t=(1+p)^t\overline{DG} \tag{13}$$

The nominal value of debt will increase with the price level. The issue of new nominal debt in every period will provide the government with precisely the revenues necessary to cover its deficit. The real value of debt DG_t/P_t will remain constant and equal to $\overline{DG}$. There has been no transfer of real wealth from the government to other sectors of the economy, even though the development of a budget deficit might give that impression to the casual observer.

Inflation gives rise here to a spurious element in the government's nominal budget balance. That spurious element does not disappear if the balance is deflated by the price level and expressed in real terms as $sg_t(=SG_t/P_t)$. Clearly sg_t will be negative whenever the nominal value, SG_t, is less than zero. The magnitude of the spurious element is readily identified. In real terms, it is given by

$$(sg_t-\overline{SG})=-p\left(\frac{DG_{t-1}}{P_t}\right) \tag{14}$$

The form of (14) suggests a correction procedure designed to purge observed values of the real government budget balance of its spurious component. The correction is a straightforward extension of the method proposed in the preceding section for the income of an individual. An inflation-adjusted value of sg_t may be obtained by simply subtracting from the observed value, the quantity $(-p_t^e(DG_{t-1}/P_t))$, where p_t^e is the expected rate of price inflation during period t, with expectation formed at the end of period t−1.[16] The appropriate measure of DG_{t-1} is the market value of debt outstanding at the end of period t − 1.

---

[15] The measured values of real *GNP* may contain spurious elements in an open economy, if that economy is a net debtor (creditor) internationally. An increase in expected inflation will cause a spurious decline (increase) in the measured value of real *GNP* of a net debtor (creditor) economy. The effect operates via the current account of the balance of payments. Measured imports include interest paid abroad, while measured exports include interest received from abroad. The real value of the current account balance of a debtor (creditor) country will show a spurious worsening (improvement in response to an increase in the expected rate of inflation.

[16] A logical question at this point is, why not compute an inflation-adjusted government balance using the actual rate of inflation p_t instead of the expected rate p_t^e? This would be simple to compute since p_t can be observed directly. The answer to this question hinges on what is considered to be the most useful concept of income. A correction based upon p_t effectively measures interest income using the *ex post* after-tax real rate of interest. A correction based upon p_t^e effectively measures interest income using the *ex ante* expected value of the after-tax real rate of interest. The latter concept has an economic interpretation; it is Hicks income, or in more familiar terms, it is the permanent income derived from fixed-income assets. The former concept has no ready interpretation. Since permanent income has been found to be a useful concept in consumer theory, the adjustment based on p_t^e is conceptually more appealing.

A final consideration in regard to the government sector applies to the effect of inflation on the measured value of real tax receipts. The pre- and postinflation values of real taxes are given by

$$\overline{TX}=\tau\left(\overline{Y}-\overline{I}+i(0)\overline{DG}\right) \tag{15a}$$

$$\frac{TX_t}{P_t}=\overline{TX}+\left(\frac{\tau}{1-\tau}\right)p\left(\frac{DG_{t-1}}{P_t}\right) \tag{15b}$$

There is a spurious effect here which makes the real tax burden faced by the private sectors of the economy appear to rise with the expected rate of inflation. Under inflation neutrality the holders of government debt have been fully compensated for the rise in p via the inflation premium in i. Their real tax burden cannot have truly increased. The term involving p on the right-hand side of (15b) is a spurious element and suggests an obvious correction factor to apply to real world observations.

B. *The Corporate Sector*

The real market value of corporations V is the present value of future after-tax real interest and dividend payments, discounted by r. Under inflation neutrality, the value of V is invariant to changes in the rate of inflation. In the economy under consideration, V is given by

$$V_t=\overline{V}=\frac{(1-\tau)(\overline{Y}-\overline{W}-\overline{I})}{r} \tag{16}$$

where $\overline{W}$ denotes the inflation-invariant real value of wages and salaries paid by corporations.

The value V is distributed between debt and equity, on some basis determined by the equity owners of corporations. Let DC_t and E_t denote the nominal values of corporate debt and equity, respectively. Assume that prior to the onset of inflation, the debt-equity ratio was constant at the value $\overline{DC/E}$. I have made no assumptions which would lead to any preferred value for this ratio, so it is reasonable to suppose that whatever factors lead to the value $\overline{DC/E}$ prior to inflation remain in force after inflation begins.[17] This implies that both DC_t and E_t will rise at the rate of inflation, so that $DC_t/E_t=\overline{DC/E}$.

Under this assumption, the real value of pretax profits Π, before and after inflation, may be compared.

$$\overline{\Pi}=\overline{Y}-\overline{W}-\overline{I}-i(0)\overline{DC} \tag{17a}$$

$$\overline{\Pi}_t=\overline{\Pi}-\left(\frac{p}{1-\tau}\right)\left(\frac{DC_{t-1}}{P_t}\right) \tag{17b}$$

There is a reduction in conventionally measured real profits as a consequence of the acceleration in inflation. This must be a spurious effect because the value of V is unaffected by inflation and the debt-equity ratio has been assumed constant. The spurious decline occurs because corporations are net debtors and inflation increases interest payments disproportionate to its effects on the price level. As was the case in my previous findings, the magnitude of the spurious element varies with p and is independent of the period of time that the economy has been inflating.

Movements in aggregate *U.S.* profit rates since the advent of high inflation in the late 1960's and early 1970's have been a topic of some debate in the recent literature. Nordhaus initiated the discussion with the observation that a postwar downward trend in profit rates seemingly accelerated during the early 1970's. I have little to say about the earlier years, but the profit performance of the last decade surely reflects in part a spurious downward tendency eminating from accelerated inflation expectations. How much of the 1970's downtrend in the profit rate can be explained by this effect is an empirical issue I leave for future research.

---

[17] Capital gains on equities are not taxed in my hypothetical economy, but full integration of the corporate and personal tax systems make it impossible for equity owners to avoid being taxed on the corporate earnings which give rise to capital gains. The tax structure does not therefore give preference to any particular debt-equity ratio. In fact, debt and equity claims are perfect substitutes because there is also no uncertainty.

The pre- and postinflation values of nominal dividends Z in my hypothetical economy are given by

$$\bar{Z}=(1-\tau)\bar{\Pi} \tag{18a}$$

$$Z_t=(1+p)^t(1-\tau)\Pi_t + (DC_t-DC_{t-1})=(1+p)^t\bar{Z} \tag{18b}$$

Corporations are able to maintain nominal dividend payments which rise proportionately with the price level after the onset of inflation. They are able to do this, even though measured profits are reduced by inflation, because the receipts derived from new nominal debt issues are exactly sufficient to compensate for the reduction in profits. This finding simply confirms that a neutral inflation does not redistribute real wealth from equity to debt holders or vice versa.

C. *The Personal Sector*

A spurious element emerges in the postinflation measured values of aggregate disposable income and personal saving in this economy in a form completely analogous to that which appeared in Section I for an individual. The onset of inflation at the rate p causes the nominal values of both measured disposable income YD_t and personal saving SP_t to rise by the amount

$$D_t=p(DG_{t-1}+DC_{t-1}) \tag{19}$$

The real values of measured income and saving rise by the spurious amount $d_t(=D_t/P_t)$, analogous to equation (7) in Section I.

The spurious element arises here because of the personal sector's status as net creditor of government and corporate debt obligations. It is interesting to note that corporate debt contributes to this effect even though ownership of corporate equity rests with the personal sector. The spurious increase in measured interest incomes of corporate debt holders is not offset by a decline in the measured incomes of equity holders. An offset does not occur because debt holders "save" their spurious income gains by acquiring new nominal debt in sufficient quantities to enable corporations to maintain a constant flow of real dividends to equity holders.

The postinflation measured value of the personal saving rate sp_t reflects the spurious element associated with income and saving.

$$sp_t=\frac{SP_t}{YD_t}=\frac{(\overline{SP}+d_t)}{(\overline{YD}+d_t)} \tag{20}$$

The form of (20) suggests an inflation-adjustment procedure designed to purge observed values of sp_t of its spurious elements. An inflation-adjusted personal saving rate derived from real world observations may be computed as

$$sp_t^*=\frac{(SP_t-p_t^eNA_{1,t-1})}{(YD_t-p_t^eNA_{2,t-1})} \tag{21}$$

where p_t^e is the one-period expected rate of price inflation and $NA_{1,t-1}$ and $NA_{2,t-1}$ are market values of net interest-bearing assets.

It is necessary in (21) to distinguish two net asset concepts in order to conform with National Income and Product Accounts (*NIP*) conventions. The term NA_1 denotes total net interest-bearing assets of the personal sector and is the appropriate concept to use in purging the spurious element from observed values of *NIP* personal saving. The term NA_2 is defined as NA_1, plus total consumer credit liabilities of the personal sector, and is the appropriate concept to apply to observed values for *NIP* disposable income. Interest payments on consumer credit liabilities are not deducted from income receipts in the computation of *NIP* values for personal income. These expenses are, however, considered to be expenses in the computation of *NIP* values for personal saving.

It is important to give the correct interpretation to the inflation-adjusted saving rate. The term sp_t^* is an estimate of the personal saving rate which would have prevailed during period t had consumers faced all of the same constraints which actually prevailed during that period, with p_t^e equal to zero. It is incorrect to interpret sp_t^* as simply the saving rate which is associated with zero expected price inflation. The latter interpretation might be appropriate for an inflation-neutral world, but in a nonneutral world, values for sp_t^* will vary over time because the values of net worth and the real rate of interest are not constant. Any period-to-period variations which are observed in sp_t^* will reflect true behavioral responses to the determinants of real saving decisions. Thus, the inflation adjusted, as opposed to the conventionally, measured value of the saving rate is the appropriate variable to use in an empirical study of saving behavior.

III. Conclusions: The *U.S.* Personal Saving Rate

The preceding analysis has revealed that spurious elements arise in many measured real variables when the *ex ante* real and nominal after-tax rates of interest diverge. This occurs whenever the expected rate of price inflation is nonzero and is independent of whether nominal rates incorporate a fully tax-adjusted inflation premium. The finding has important implications both for the intertemporal comparison of economic variables and for empirical research. If economists are unaware of the existence of spurious elements, movements in observed variables which occur during periods of accelerating or decelerating inflation are apt to be puzzling. They may be tempted to conclude that traditional economic theory is inadequate to explain observed phenomena. The remedy is to identify the spurious components and compute inflation-adjusted values for the affected variables. A number of suggestions have been made as to how this might be done.

In order to give some indication of the magnitudes involved, I will forego the usual concluding comments and end this paper with an empirical computation of an inflation-adjusted time-series for the *U.S.* personal saving rate. This is the aggregate variable which is probably more susceptible to inflation-related spurious movements than any other. It is also the variable which has prompted the greatest volume of empirical literature over the past decade.

Following a long period of relative stability, the *U.S.* personal saving rate began to rise in the late 1960's. It reached a postwar high of 7.8 percent in 1973, before declining to more "normal" values in 1976–78. Preliminary estimates for 1979 suggest a measured value which is low by historical standards. Our computations will show that on an inflation-adjusted basis the value of the *U.S.* saving rate has been well below the postwar average since 1976.

A number of research efforts have been aimed at explaining the behavior of the *U.S* saving rate during the early 1970's. Most studies (see for example, Thomas Juster and Paul Wachtel, 1972a, b) have attributed this phenomenon to increases in inflation risk.[18] Theoretically, households ought to alter their consumption-saving plans in response to inflation risk. However, theory does not predict the direction of the effect on saving: the saving rate could rise or fall as risk increases, depending upon individual's preferences and the distribution of their lifetime earnings. Empirical studies which include some measure of inflation risk among the explanatory variables in a regression with the measured saving rate as dependent variable are very likely to obtain a positive coefficient. The reason for this is that measures of inflation risk tend to rise and fall with the level of inflation. Attitudes towards risk appear to "explain" the behavior of saving in the early 1970's, but this may be no more than a spurious association due to an inappropriately measured dependent variable.

Inflation-adjusted values for the *U.S.* personal saving rate were computed following equation (21) on an annual basis for the period 1958–78. These values appear as sp_t^* in Table 1, along with values for the measured saving rate and estimates of one-year expected rate of price inflation over this interval. Details of the computations appear in the Appendix.

---

[18] One exception has been a study by Angus Deaton. He offers an argument why *unexpected* inflation might lead to an increase in saving, above and beyond its impacts via lowered real wealth and real income, but does not present an explanation of why expected inflation might affect saving.

TABLE 1—INFLATION-ADJUSTED VALUES FOR THE *U.S.* PERSONAL SAVING RATE
(All values measured as percentages)

	Measured Rate sp	Inflation-Adjusted Rate sp^*	Expected Rate of Inflation p^e
1958	6.8	5.9	2.3
1959	5.6	4.6	2.3
1960	4.9	4.0	2.3
1961	5.6	4.7	2.1
1962	5.3	4.6	1.5
1963	4.7	3.9	1.5
1964	6.0	5.3	1.5
1965	6.4	5.8	1.4
1966	6.5	5.8	1.6
1967	7.5	6.7	2.3
1968	6.5	5.5	2.4
1969	5.6	4.2	3.5
1970	7.4	5.9	4.0
1971	7.7	6.0	4.5
1972	6.2	4.1	4.4
1973	7.8	6.0	3.9
1974	7.3	5.5	4.9
1975	7.7	4.4	9.7
1976	5.8	2.8	7.3
1977	5.0	1.8	6.9
1978	4.9	2.2	6.9

The upward drift apparent in the measured saving rate from 1968–75 is absent in the inflation-adjusted series. Values for *sp** show cyclical movements over this subperiod, but are no different on average from values computed for the earlier period. The maximum effect of the spurious element on the measured saving rate occurs in 1975 with the peak value for expected inflation. The spurious element is estimated to have been 3.3 percentage points during 1975, or more than 40 percent of the observed saving rate.

The value of *sp** shows a sharp decline in 1976 which is sustained over the remainder of the interval. Whether this represents some behavioral phenomenon or simply reflects an error in the adjustment procedure is a question for future research. However, it is worth noting that the measured saving rate for the United States has declined relative to that in other industrialized countries since 1975. When a similar adjustment is made to Canadian data, the values of *sp** for the late 1970's look very similar to values for earlier periods. One is therefore tempted to view the recent *U.S.* experience as a behavioral response unique to the *U.S.* economy.

It would be premature to draw further conclusions from these results, but it does seem clear that spurious elements can account for most of the anomalies in observed saving behavior during the first half of the 1970's. The magnitudes involved are substantial. There is every reason to suppose that inflation adjustment of other income and saving variables would also reveal significant spurious elements.

APPENDIX: INFLATION-ADJUSTMENT COMPUTATIONS

The annual series for inflation-adjusted rates of personal saving was computed for 1958–78 from

$$\text{(A1)}\quad sp_t^* = \frac{SP_t - p_t^e NA_{t-1}}{YD_t - p_t(NA_{t-1} + CC_{t-1})}$$

where SP is measured personal saving, YD is measured disposable income, p^e is the one-year expected rate of inflation, NA is the end-of-period market value of net fixed-income assets, and CC is the market value of consumer credit. Values for SP and YD were taken directly from the *NIP* accounts. Values for p^e, NA, and CC were constructed.

The obvious source of data for net assets of persons is the household sector of the Flow of Funds Accounts (*FF*) published in the *Federal Reserve Bulletin*. The *FF* data are recorded at book values, and some method for estimating market values had to be derived. The market value A_t of an interest-bearing asset is related to its book value B_t, via

$$\text{(A2)}\quad A_t = \frac{c}{i_t}B_t + B_t\left(1 - \frac{c}{i_t}\right)\left(\frac{1}{1+(1-\tau)i_t}\right)^n$$

where c is the coupon rate of interest payable on the asset, n is its remaining term to maturity, i_t is the nominal rate of interest at time t, and τ is the effective tax rate on earnings from the asset. For large values of

n,

$$A_t \doteq \frac{c}{i_t} B_t \tag{A3}$$

Approximation (A3) was used to convert *FF* Accounts book values of household assets to market values. Values for c are not directly observable, but since coupon rates typically equal the market rate of interest at the time an asset is first issued, the effective value of c prevailing on total household net assets at time t is some weighted average of the current and past values of i.

$$c_t = \sum_{j=0}^{K} w_j i_{t-j};\ \sum w_j = 1.0 \tag{A4}$$

The weight w_j represents the share of outstanding net assets acquired in period $t-j$.

The market value of net fixed-income assets of the household sector was estimated as

$$NA_t = \frac{NAB_t}{i_t} \sum_{j=0}^{K} w_j i_{t-j} \tag{A5}$$

where *NAB* represents the end-of-year book values of household net assets from the *FF* Accounts.

In performing these computations, i_t was measured by the yield on long-term *U.S.* government bonds and the summation term was based on a simple four-year average with $w_j = 1/4$ for $j = 0, 1, 2, 3$. An analogous computation was performed to estimate the market value of consumer credit.

There is little to justify the use of the long-term bond yield other than the fact that it is available over a longer historical period than most other nominal interest rates. Computations based on this yield could be made for the later 1950's. Alternative computations of sp^*, based on the shorter time series for the yield on nine-to twelve-month *U.S.* Treasury bills, were made for the period 1964–78. The results did not differ appreciably over this interval from the values of sp^* based on the long-term bond yield.

The use of a simple four-year average of values of i_t to approximate the coupon rate was an arbitrary assumption. Coupon rates obviously vary from one type of asset to another, and setting the value of K in (A5) between two and six years seems a priori reasonable. The value $K=3$ represents a midrange compromise. Computations of sp^* based on alternative values for K within this range did not differ appreciably from the reported series.

The *FF* Accounts definition of the household sector differs conceptually from the *NIP* Accounts definition of the personal sector. The former regards assets of government pension funds as household assets while the latter treats flows in and out of these funds as transfers between persons and governments. In order to use the net asset concept compatible with *NIP* accounts measures of disposable income and personal saving, the book value *NAB* used in (5) was computed as net fixed-income assets from the *FF* Accounts household sector, less household holdings of public pension assets.

Inflation expectations were estimated from standard time-series analysis, based on the proposition that in an efficient market all useful information pertaining to future inflation should be contained in observations on past rates of inflation. After some preliminary examination of the data to determine stationarity, the following specification was deemed best to describe the time-series properties of observed inflation rates.

(A6)

$$(p_t - p_{t-1}) = \sum_{j=1}^{T} \alpha_j (p_{t-j-1} - p_{t-j-2}) + \varepsilon_t$$

The variable ε_t is a white noise disturbance and the summation term represents the portion of any change in inflation between periods $t-1$ and t which is predictable on the basis of past observations.

Observations on p_t were obtained from annual data on the rate of change in the implicit price deflator for consumer expenditures from the *NIP* accounts for the period 1947–78. The equation was first estimated using ordinary least squares for the period 1952–78 to determine the optimal value of T. A value $T=3$ gave maximum $\bar{R}^2$ over this interval and was used in all remaining regressions.

The equation was then estimated for 1952–57, and the estimated parameters used to predict the value of inflation for 1958. Actual data for 1958 were added to the sample and the equation was reestimated for 1952–58. Results of this regression were used to make a prediction for 1959. The process was repeated in step-wise fashion until predicted values were obtained for each year in the 1958–78 interval. These predicted values were used as estimates of p_t^e, the one-year rate of price inflation expected to prevail during period t. Values for p_t^e appear in Table 1.

REFERENCES

J. L. Carr and L. B. Smith, "Money Supply, Interest Rates, and the Yield Curve," *J. Money, Credit, Banking*, Aug. 1972, *4*, 582–94.

______, J. E. Pesando, and L. B. Smith, "Tax Effects, Price Expectations and the Nominal Rate of Interest," *Econ. Inquiry*, June 1976, *14*, 259–69.

M. R. Darby, "The Financial and Tax Effects of Monetary Policy on Interest Rates," *Econ. Inquiry*, June 1975, *13*, 226–76.

A. Deaton, "Involuntary Saving Through Unanticipated Inflation," *Amer. Econ. Rev.* Dec. 1977, *67*, 899–910.

O. Eckstein and M. Feldstein, "The Fundamental Determinants of the Interest Rate," *Rev. Econ. Statist.*, Nov. 1970, *52*, 363–75.

M. Feldstein, "Inflation, Income Taxes and the Rate of Interest," *Amer. Econ. Rev.*, Dec. 1976, *66*, 809–20.

______ and L. Summers, "Is the Rate of Profit Falling?," *Brookings Papers*, Washington 1977, *1*, 211–28.

______ and ______, "Inflation, Tax Rules, and the Long-term Interest Rate," *Brookings Papers*, Washington 1978, *1*, 61–99.

Irving Fischer, "Appreciation and Interest," *Publications of the American Economic Association*, Aug. 1896, *11*, 331–442.

______, *The Theory of Interest*, New York 1930.

Milton Friedman, *A Theory of the Consumption Function*, Princeton 1957.

John R. Hicks, *Value and Capital*, Oxford 1939.

G. P. Jenkins, "Inflation, Its Financial Impact on Business in Canada," disc. paper no. 72, Economic Council of Canada, Ottawa 1977.

F. T. Juster and P. Wachtel, (1972a) "Inflation and the Consumer," *Brookings Papers*, Washington 1972, *1*, 71–121.

______ and ______, (1972b) "A Note on Inflation and the Saving Rate," *Brookings Papers*, Washington 1972, *3*, 765–78.

W. D. Nordhaus, "The Falling Share of Profits," *Brookings Papers*, Washington 1974, *1*, 169–208.

James E. Pesando and Samuel A. Rea, Jr., *Public and Private Pensions in Canada: An Economic Analysis*, Ontario Econ. Council Res. Stud., Toronto 1977.

T. J. Sargent, "Interest Rates and Expected Inflation: A Selective Summary of Recent Research," *Explor. Econ. Res.*, Summer 1976, *3*, 303–25.

J. B. Shoven and J. I. Bulow, "Inflation Accounting and Non-Financial Corporate Profits: Financial Assets and Utilities," *Brookings Papers*, Washington 1976, *1*, 15–66.

P. Wachtel, "Inflation, Uncertainty, and Saving Behavior," *Explor. Econ. Res.*, Fall 1977, *4*, 558–78.

W. P. Yohe and D. S. Karnosky, "Interest Rates and Price Level Changes, 1952–69," *Federal Res. Bank St. Louis Rev.*, Dec. 1969, *51*, 18–38.

Board of Governors of the Federal Reserve System, *Fed. Reserve Bull.*, various issues.

Two Illustrations of the Quantity Theory of Money

By ROBERT E. LUCAS, JR.*

This paper presents empirical illustrations of two central implications of the quantity theory of money: that a given change in the rate of change in the quantity of money induces (i) an equal change in the rate of price inflation; and (ii) an equal change in nominal rates of interest. The illustrations were obtained by comparing moving averages of the three variables in question, using quarterly *U.S.* time-series for the period 1953–77. Readers may find the results of interest as additional confirmation of the quantity theory, as an example of one way in which the quantity-theoretic relationships can be uncovered via atheoretical methods from time-series which are subject to a variety of other forces, or as a measure of the extent to which the inflation and interest rate experience of the postwar period can be understood in terms of purely classical, monetary forces.

The theoretical background of the study is reviewed, very briefly as it is familiar material, in the next section. The data processing methods are described and rationalized in Section II. The illustrations resulting from the application of these methods are in Section III. Section IV contains some decompositions of postwar time-series and concluding comments.

I. Theoretical and Empirical Background

The two quantity-theoretic propositions stated in the introduction possess a combination of theoretical coherence and empirical verification shared by no other propositions in monetary economics. By "theoretical coherence," I mean that each of these laws appears as a characteristic of solutions to explicit theoretical models of idealized economies, models which give some guidance as to why one might expect them to obtain in reality, also as to conditions under which one might expect them to break down. For present purposes, Miguel Sidrauski's monetary version of the Solow-Swan one-sector model of economic growth (1967a, b) is perhaps the most useful, single theoretical illustration. In that model, both laws appear as explicit, necessary characteristics of the stationary solution of the differential equations which describe equilibrium in the system. To restate this in a way which is more suggestive empirically, they appear as characteristics of long-run average behavior in the model economy.

Both of these laws are, as is clear from the Sidrauski example, propositions about the consequences of a unit's change. Thus neither appears to depend crucially on particular features of the preferences and technology postulated by Sidrauski. It is not difficult to construct other examples to illustrate the insensitivity of these laws to variations in the structure of the economy. In particular, if stochastic elements are introduced, the laws are reinterpreted to apply to means of theoretical stationary distributions or, as before, to long-run average behavior.[1]

Sidrauski's example, together with variations appearing in the literature both before and since he wrote, also suggests some qualifications or limitations to these laws.

---

*Professor of economics, University of Chicago. I have benefited from discussions with Jacob Frenkel, Sherwin Rosen, Thomas Sargent, Gary Skoog, and William Wecker. I would also like to thank Frederick Miller and Christopher Flinn for their assistance, and the National Science Foundation for its support.

[1] This interpretation of the quantity theory of money as a set of predictions about the long-run average behavior of a general equilibrium system is different from, though not inconsistent with, Milton Friedman. There, Friedman stresses the stability of the market demand function for money, a property which is neither necessary nor sufficient for the quantity theory to obtain in the sense used here.

First, Sidrauski's version of the neoclassical model does not exhibit the Mundell-Tobin effect of a monetary expansion: the possibility that an inflation, by reducing the real yield on money, will shift saving to real capital accumulation. If this effect is important, it would force us to modify the second law to predict interest rate increases by *less* than the increase in the monetary growth rate (due to the decline in the real return on capital, offsetting the inflation premium).[2] Theoretically, I think it is clear from related work (see, for example, David Levhari and Don Patinkin, Stanley Fischer, and Ronald Michener) that only a very coincidental combination of assumptions produces an absence of a Mundell-Tobin effect in Sidrauski's example and that, in general, one does not want to view this effect as ruled out on prior, logical grounds. This conclusion, of course, leaves us free to hope that the required modifications are minor enough to be neglected in some applications.

Second, and perhaps more fundamental, theory at this level gives no guidance as to the measurement of the quantity of money, or as to which (if any) of the available time-series on monetary aggregates corresponds to the variable theoretically termed "money." (Of course, it also gives no guidance as to the empirical definition of "the price level," but there is a good deal of other economic theory which does.) As recent theoretical work of John Bryant and Neil Wallace and Marco Martins has emphasized, this question of which monetary aggregate one would theoretically *expect* to move in proportion to prices is much more open than has traditionally been recognized. In the experiments reported below, money means $M1$, but the arbitrariness of this measurement choice should be emphasized at the outset, particularly as it is likely that very similar results would have been obtained under a variety of other choices.

In summary, then, we have specific theoretical examples exhibiting both quantity-theoretic laws in clear, exact form, and others which suggest possibly important qualifications. This is all we can ever hope for from our theory: some strong clues as to what to look for in the data; some warnings as to potential sources of error in these predictions. This is the theoretical coherence of the neoclassical laws.

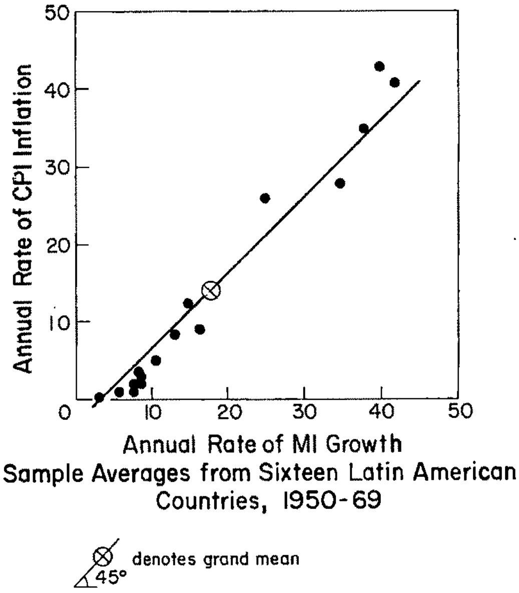


FIGURE 1

Since the two quantity-theoretic laws are obtained as characteristics of steady states, or limiting distributions, of theoretical models, the ideal experiment for testing them would be a comparison of long-term average behavior across economies with different monetary policies but similar in other respects. Many such tests of the first law are available;[3] a particularly clean example is shown in Figure 1. These data are taken from Robert Vogel's study of inflation in sixteen Latin American economies, using annual data for the period 1950–69.[4] Vogel does not report the interest rate data which would have permitted a comparable test of the Fisherian interest-inflation relationship. In general, such evidence is difficult to obtain, no doubt due to the fact that in inflationary economies published interest rates are rarely left free to reach their equilibrium levels.

---

[2] Since interest payments are taxable, the maintenance of a given real yield on bonds would require interest rates to rise by more than the inflation rate. This effect will offset, and perhaps even reverse, the Mundell-Tobin effect.

[3] See in particular Anna Schwartz.

[4] The countries included in Vogel's study are Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, Peru, Uruguay, and Venezuela.

The line in Figure 1 is drawn through the grand mean of the $16 \times 20 = 320$ annual money growth rate inflation pairs in Vogel's sample. This is the one "free parameter" permitted by the theory. Its slope is 45°, as specified theoretically: it is not fit to the data. It is hard to imagine a nonvacuous economic prediction obtaining stronger confirmation than that shown in Figure 1. This is the kind of "empirical verification" of the quantity theory on which economists who assign it a central theoretical role base most of their confidence.

In the absence of the kind of decisive natural experiment used by Vogel, one could in principle test the neoclassical laws by deriving their implications for the parameters of a structural econometric model. This course, while attractive in theory (since it broadens considerably the class of data which might shed light on the laws), is in practice a difficult one, since it involves nesting the two hypotheses in question within a complex maintained hypothesis, which must be accepted as valid in order to carry out the test. The virtue of relatively atheoretical tests, such as carried out by Vogel, is that they correspond to our theoretically based intuition that the quantity theoretic laws are consistent with a wide variety of possible structures. If so, it would be desirable to test them independently and then, if confirmed, to *impose* them in constructing particular structural models, rather than to proceed in the reverse direction. It would be of value, then, to have measurement techniques which are atheoretical in the sense of Vogel's but which can be applied to continuous time-series for a single economy. The use of one such technique is illustrated below.

II. Data and Data Processing Methods

The time-series used in this study are the money supply ($M1_t$), the consumer price level (P_t) and the ninety-day Treasury bill rate (r_t). The value of $M1_t$ for quarter t is demand deposits plus currency outside banks, for the first month of the quarter, seasonally adjusted, taken from successive issues of the *Federal Reserve Bulletin*. The *CPI* is similarly timed, not seasonally adjusted, from the Consumer Price Index. The bill rate is that used and described by Eugene Fama.

I shall work with the following transformed variables:

$$X_{0t} = ln(M1_{t+1}) - ln(M1_t)$$

$$X_{1t} = ln(P_{t+1}) - ln(P_t)$$

$$X_{2t} = r_t$$

Scatter diagrams of X_{1t} and X_{2t} against X_{0t} are given in Figures 2 and 3, in the next section. These figures seem to capture fairly well what people mean when they say that the quantity theory of money is not a "short-run" relationship.

The general idea of what follows will be to examine scatter diagrams of $X_{it}(\beta)$, $i=1$, 2, against $X_{0t}(\beta)$ where for $i=0, 1, 2$, $X_{it}(\beta)$ is the two-sided exponentially weighted moving average given by[5]

$$(1) \qquad X_{it}(\beta) = \alpha \sum_{k=-\infty}^{\infty} \beta^{|k|} X_{i,t+k}$$

where $$\alpha = \frac{1-\beta}{1+\beta}, \; 0 \leqslant \beta < 1$$

The effect of the filter (1) is to smooth the original series; indeed, as β approaches unity, the filtered observations $X_{it}(\beta)$ approach the sample average values of the original series. In the latter case, plots of $X_{1t}(\beta)$ and $X_{2t}(\beta)$ against $X_{0t}(\beta)$ will degenerate to a point, vacuously lying on a line with slope 45°. Our interest will be in seeing whether the points $(X_{0t}(\beta), X_{it}(\beta))$, $i=1, 2$, fall on a 45° line for β-values less than unity, providing a time-series confirmation of the cross-country results obtained by Vogel and others. Viewed as a measurement procedure, the test of this method will be the quality of the pictures it yields. It may be useful first, however, to look in more detail into what the filter (1) does to a time-series, and what statistical and economic rationales may underlie its use.

---

[5] Here and below I write as though the entire doubly infinite record were available for each variable. In the calculations, the algorithm described by Thomas Cooley, Barr Rosenberg, and Kent Wall was used. This algorithm permits the assignment of a diffuse prior on X_{it} values outside the sample period which appear in the doubly infinite sum (1). With beliefs about points prior to 1953 and after 1977 so described, it calculates posterior means of the slowly moving "signal," called s_t below. Except for points near the beginning and the end of the sample period, virtually identical results were obtained simply by replacing missing observations in (1) by zeros. So as not to present results which are unduly dependent on the way out-of-sample X_{it} values are treated, numbers for 1953-54 and 1976-77 are not plotted.

The Fourier transform of the filter given in (1), with α free, is for $0\leqslant\omega\leqslant\pi$,

$$(2)\quad f(\omega;\alpha,\beta)=\alpha\sum_{k=-\infty}^{\infty}\beta^{|k|}e^{-i\omega k}$$

$$=\frac{\alpha(1-\beta^2)}{1+\beta^2-2\beta cos(\omega)}$$

One verifies that $f(0)=1$ if $\alpha=(1-\beta)/(1+\beta)$, that $f(\pi)>0$ and that $f'(\omega)<0$ for all $0\leqslant\omega\leqslant\pi$. Also $f''(0)<0$ and $f''(\pi)>0$; $f''(\omega)$ changes sign once, at the unique ω value at which $X=cos(\omega)$ is a positive root of

$$X^2+\frac{1+\beta^2}{2\beta}X-2=0$$

For high β-values (for example, near 0.9) this root occurs very near $\omega=0$. Since the spectral density of the filtered series $X_{it}(\beta)$ is just the spectral density of X_{it} multiplied by $f(\omega;\alpha,\beta)$, one sees that the filter (1) retains power at very low frequencies, while sharply reducing power at higher frequencies.

Filters of the form (1) are solutions to a well-known signal-extraction problem, the form of which may also be instructive.[6] Let $\{v_t, w_t\}$ be a white noise process with mean (0,0) and covariance matrix $\sigma^2\begin{bmatrix}\theta & 0\\ 0 & 1\end{bmatrix}$. Define the processes u_t and s_t by

$$u_t=s_t+v_t$$

$$s_t=\rho s_{t-1}+w_t$$

$$0<\rho<1$$

Imagine that this structure, including the values of the parameters θ, σ^2, and ρ, is known and that one has observations on the u_t, $t=-\infty,\dots,\infty$. It is desired to obtain minimum variance unbiased estimators $\hat{s}_t$ of the sequence of signals s_t. Projecting s_t on u_t, $t=-\infty,\dots,\infty$

$$s_t=\hat{s}_t+\eta_t$$

$$=\sum_{k=-\infty}^{\infty}\gamma_k u_{:+k}+\eta_t$$

where $E(u_t\eta_s)=0$, all s, t. The coefficients γ_k must satisfy the normal equations:

$$E(u_{t+j}s_t)=\sum_{k=-\infty}^{\infty}\gamma_k E(u_{t+j}u_{t+k})$$

$$j=-\infty,\dots,\infty$$

Taking the Fourier transform of both sides:[7]

$$f_{us}(\omega)=f_{uu}(\omega)f_\gamma(\omega),\ 0\leqslant\omega\leqslant\pi$$

or, exploiting the particular structure of the process assumed here,

$$f_{ss}(\omega)=[f_{ss}(\omega)+f_{vv}(\omega)]f_\gamma(\omega)$$

Solving for $f_\gamma(\omega)$ gives

$$(3)\quad f_\gamma(\omega)=\frac{1}{1+\theta[1+\rho^2-2\rho cos(\omega)]}$$

since $f_{vv}(\omega)=\theta\sigma^2$ and $f_{ss}(\omega)=[1+\rho^2-2\rho\, cos(\omega)]^{-1}\sigma^2$.

---

[6] See Peter Whittle (ch. 5) for a discussion of this and other examples.

[7] For two time-series $\{x_t\}$ and $\{y_t\}$, the notation $f_{xy}(\omega)$ means

$$f_{xy}(\omega)=\sum_{k=-\infty}^{\infty}e^{-i\omega k}Cov(x_{t+k}, y_t)$$

For the functions $f(\omega;\ \alpha,\beta)$ in (2) and $f_y(\omega)$ in (3) to be the transforms of the same filter, it is necessary that the right-hand sides of these equations be identically equal in ω on $[0,\pi]$. This requires that β in (2) be that root of

$$0=1-\beta\left[\frac{1+\theta(1+\rho^2)}{\theta\rho}\right]+\beta^2$$

which lies in $(0,1)$. Given β, α must satisfy

$$\alpha=\frac{\beta}{\theta\rho(1-\beta^2)}$$

The particular filter given in (1) is a one-parameter family in which α and β are constrained by $\alpha=(1-\beta)(1+\beta)^{-1}$. This case is seen to correspond to the limiting situation where $\rho=1$ and β solves[8]

$$0=1-\left[\frac{1+2\theta}{\theta}\right]\beta+\beta^2 \tag{4}$$

Hence if the variance of the "noise" v_t is small relative to the variance of w_t $(\theta\cong 0)$, the root β of (4) in $(0,1)$ will be near 0. This means that the current observation u_t is a good estimate of the true signal s_t. In our economic application, where s_t is taken to be that part of a time-series which is dominated by quantity-theoretic forces, this would correspond to a situation in which other "real" forces play a negligible role. At the other extreme, when the noise variance is high (θ large), β will be near one, and the best estimate of the true signal at t will be a very long moving average of the observed u_t.[9]

This purely statistical rationale for experimenting with the filter (1) has no basis in economic theory, and a little reflection suggests that none will be forthcoming: a *good* economic theory accounting for both quantity-theoretic and other forces on interest rate and price series would surely suggest the use of a "sharper" filter than (1). Nevertheless, the following scenario may be helpful. Imagine an economy in which the rate of monetary growth is a constant, known to agents, plus noise. The known, constant component is incorporated exactly into inflation and interest rates, with a negligible Mundell-Tobin effect. The monetary noise induces noise in interest and prices. In this example, the signal s_t represents the "constant" known, common component in monetary growth, price inflation, and interest rates. The noise v_t will be different for the different series.

Next, imagine that s_t, while constant for long stretches of time, infrequently changes to a new value from time to time. That is, model s_t by

$$s_t=\begin{cases} s_{t-1} & \text{with} \quad prob\ \lambda \\ \tilde{s}_t & \text{with} \quad prob\ 1-\lambda \end{cases}$$

where $\tilde{s}_t$ is serially independent with mean 0, and variance τ^2, and where $1-\lambda$ is "small." This process has the same covariance structure as the "signal" used in the statistical example above, with $\rho=\lambda$ and $\sigma^2=(1-\lambda)\tau^2$.

For an econometrician to treat this economy as posing a signal processing problem of the above type, one assumes that the "structural changes" in s_t are perfectly understood by agents as they occur, but cannot be observed by the econometrician. Hence the use of a *two-sided* moving average filter.[10]

The hope in applying this filter is not, of course, that an economic model of this type holds *exactly*. It is rather the general idea that the actual series may be generated by a *very* slowly changing structure of monetary policy, with business cycle activity occurring

---

[8] This is the quadratic John Muth arrived at, for the same reasons, in his study of the permanent income hypothesis.

[9] In the application below, the noise component is not serially uncorrelated as assumed in the example just discussed. For a more general discussion of the rationales for the use of a filter such as that described by (1), see Christopher Sims.

[10] This is not, of course, a compelling reason for using a two-sided filter. It is simply the condition under which it would be optimal to do so. In general, agents *know* only the past (arguing for a one-sided backward filter) but they *care* only about the future, and probably process much more information in forecasting that part of the future relevant to their own decisions than we econometricians can observe (arguing for a one-sided forward filter).

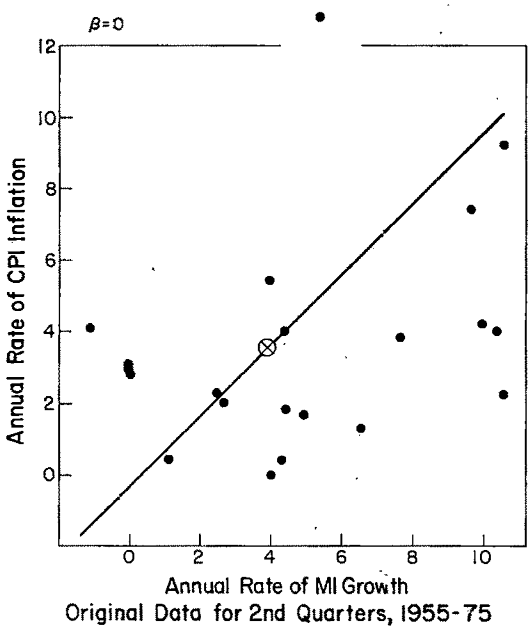


FIGURE 2

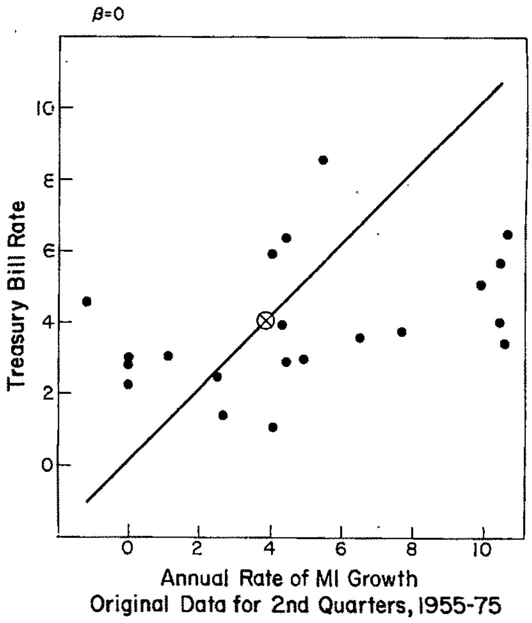


FIGURE 3

at higher frequencies superimposed. One can construct examples of economies fitting this description rather well, and one can construct other examples deviating very sharply from this description.

III. Illustrations

Figures 2 and 3 present scatter diagrams of inflation and interest rates, respectively, against rates of *M*1 growth. As remarked earlier, no relationship is evident.

Figures 4 and 5 are plots of moving averages with weights (on all series) equal to 0.5. That is, Figure 4 plots $X_{1t}(0.5)$ against $X_{0t}(0.5)$ and Figure 5 plots $X_{2t}(0.5)$ against $X_{0t}(0.5)$. Figures 6 and 7 utilize $\beta=0.8$; 8 and 9, $\beta=0.9$; and 10 and 11, $\beta=0.95$. All figures are drawn to the same scale. To avoid clutter, only points for the second quarter of each year are plotted. For high β-values, it is clear that this choice, while arbitrary, is of no consequence. Points for the first two years (1953–54) and last two (1967–77) are not plotted, though they were used in calculating the 1955–75 observations.

Given the preparatory discussion in Section II, little need be said about these figures. It is evident that a filter with $\beta=0.5$ does not quite extract the quantity-theoretic signal. A β-value of 0.9 reveals a clear 45° line, as predicted by the quantity theory and produces a picture about as clear as Vogel's cross-country estimates (Figure 1); $\beta=0.95$ is clearer still. If a Mundell-Tobin effect were present, and if it dominated tax effects, this would show up in the odd-numbered figures as a line with slope less than 45°. Perhaps this may be seen, for example, in Figure 9. Since deviations of the moving averages from the 45° line are sure to exhibit patterns, the temptation to read Figure 9 (or 11) this way should probably be resisted.

It should be added that subjecting any two series to moving-average filtering of the type used here will cause a "pattern" of some kind to emerge. To illustrate, Figure 13 plots a two-sided moving average of the unemployment rate,[11] with $\beta=0.9$, against

---

[11] Last month of quarter, not seasonally adjusted, from *Employment and Earnings*.

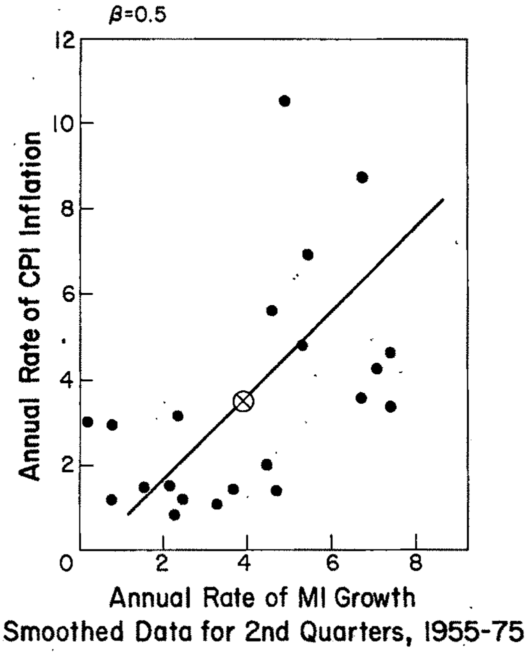


FIGURE 4

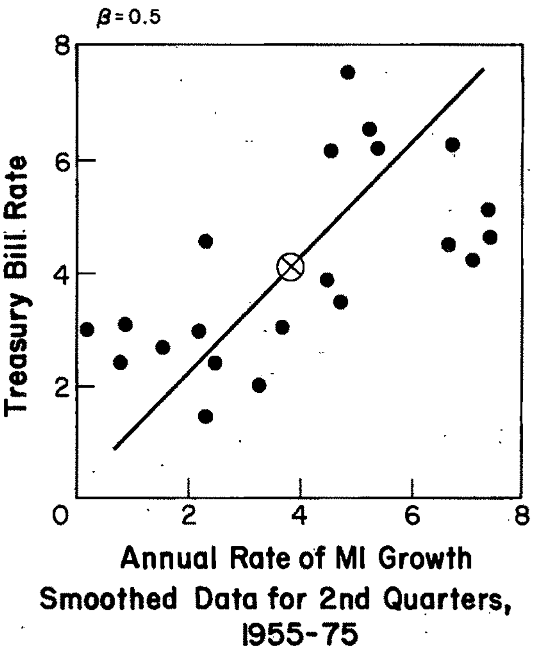


FIGURE 5

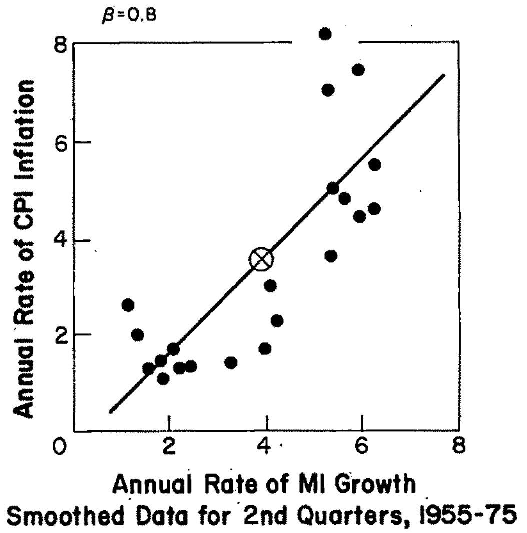


FIGURE 6

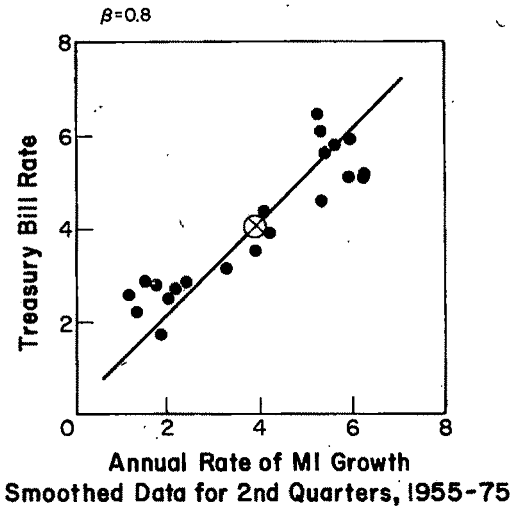


FIGURE 7

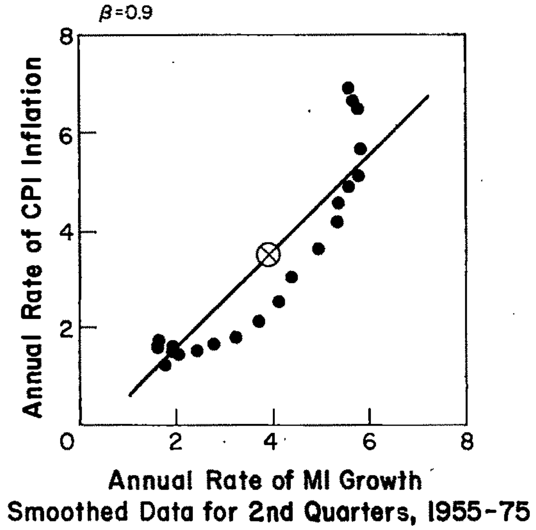


FIGURE 8

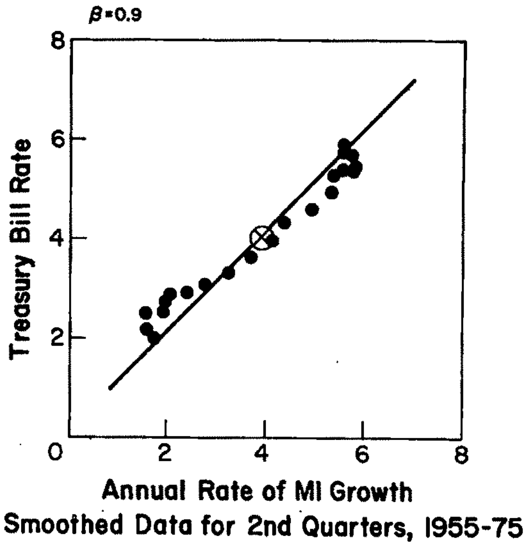


FIGURE 9

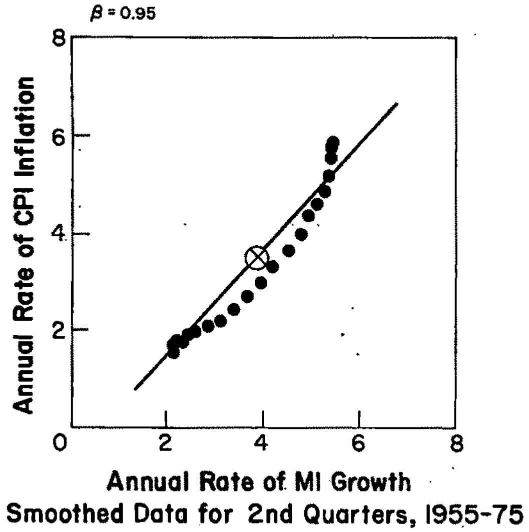


FIGURE 10

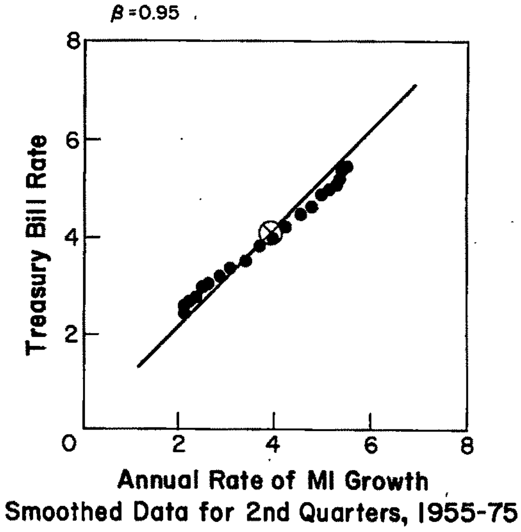


FIGURE 11

the smoothed monetary change $X_{0t}(0.9)$, while Figure 12 plots one raw variable against the other. Again, one sees order of a sort emerging from confusion but it is an order that makes no sense economically. The difference between this order and that displayed in Figures 8 and 9 is that the latter is an implication of a coherent economic theory.

Since the comparison of $X_{it}(.9)$, $i=1, 2$, to $X_{0t}(.9)$ in Figures 8 and 9 utilizes only low-frequency components of the original series, these figures will illustrate the quantity theory well only if the time-series used convey information on low-frequency movements in X_{0t}. In the absence of such information, the method applied above will produce merely a "blob" at the sample means of

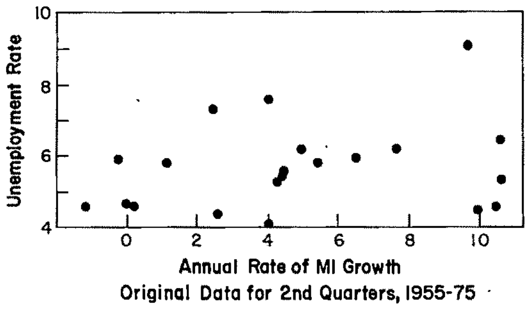


FIGURE 12

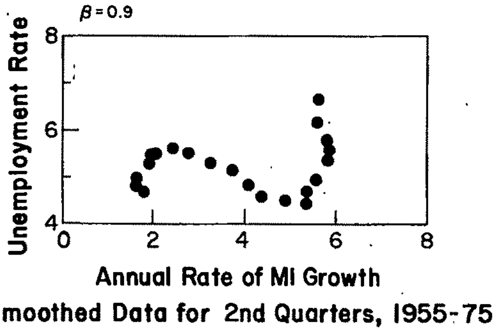


FIGURE 13

$X_{it}(\beta)$, $i=1$, 2, and $X_{0t}(\beta)$ as β approaches unity, even if the quantity theory is valid. This is the time-series equivalent of the observation that if the countries studied by Vogel had had similar rates of monetary growth over his sample period, his method would not have produced a clear 45° line. That is, these methods will yield clear results only if a good enough "experiment" has been run by "nature" over the sample period used.

IV. Concluding Remarks

The filtering techniques described and applied in Sections II and III represent what might be called a "minimal" use of the quantity theory of money, in the sense that they utilize only the widely agreed-upon "long-run" implications of that theory. To this was added the hunch that identifying long-run with "very low frequency" might isolate those movements in postwar inflation and interest rates which can be accounted for on purely quantity-theoretic grounds. Figures 8 and 9 (or 10 and 11) confirm both the hunch and the underlying theory.

Figures 14 and 15 plot actual postwar inflation and interest rates, respectively, against time (i.e., X_{1t} and X_{2t}). On each diagram is also plotted the corresponding series with the smoothed portion subtracted (that is, $X_{1t}-X_{1t}(0.9)$ and $X_{2t}-X_{2t}(0.9)$). Evidently, both the inflation and the high interest rates of the 1970's are well accounted for by the quantity theory or, to put the same point backwards, any nonmonetary explanation of these trends would lead to

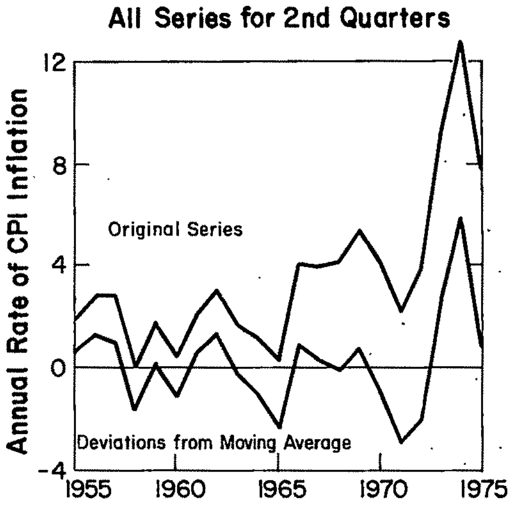


FIGURE 14

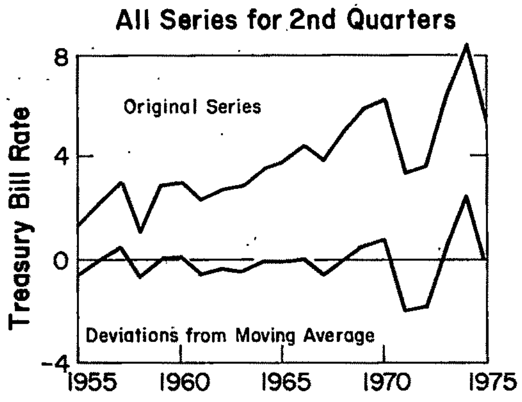


FIGURE 15

large, unexplained deviations from the relationships depicted so clearly in Figures 8–11.

The method applied in this paper involves decomposing movements in money and other nominal variables into two components, one of which I have called quantity theoretic and the other of which has been left unlabeled. This raises the question of the relationship of this decomposition to the clearly related decompositions of Thomas Sargent, and Robert Barro, among others, of monetary movements into "anticipated" and "unanticipated" components. Though it would be hard to spell out the details, my opinion would be that *all* of what I have called $X_{0t}(0.9)$ should be identified as anticipated in the Sargent-Barro sense, and *in addition*, that much of my $X_{0t}-X_{0t}(0.9)$ should also be thought of as anticipated. Indeed, this is what I mean by referring to the methods above as a *minimal* use of the quantity theory.

Putting the matter in this way should make it clear that no one decomposition method can dominate the other. By using weaker theory, one is more confident that his filter has not incorrectly labeled noise as signal; on the other side, there is no doubt that the methods used in this paper have not fully extracted from the series all that the quantity theory can account for.

REFERENCES

R. J. Barro, "Unanticipated Money Growth and Unemployment in the United States," *Amer. Econ. Rev.*, Mar. 1977, *67*, 101–15.

J. Bryant and N. Wallace, "The Inefficiency of Interest-Bearing National Debt," *J. Polit. Econ.*, Apr. 1979, *87*, 365–81.

T. F. Cooley, B. Rosenberg, and K. D. Wall, "A Note on Optimal Smoothing for Time-Varying Coefficient Problems," *Annals Econ. Soc. Measure.*, Fall 1977, *6*, 453–56.

E. Fama, "Short-Term Interest Rates as Predictors of Inflation," *Amer. Econ. Rev.*, June 1975, *65*, 269–82.

S. Fischer, "Capital Accumulation on the Transition Path in a Monetary Optimizing Model," work. paper, M.I.T., Feb. 1978.

Milton Friedman, "The Quantity Theory of Money—A Restatement," in his *Studies in the Quantity Theory of Money*, Chicago 1956.

D. Levhari and D. Patinkin, "The Role of Money in a Simple Growth Model," *Amer. Econ. Rev.*, Sept. 1968, *58*, 713–53.

M. A. C. Martins, "A Nominal Theory of the Nominal Rate of Interest and the Price Level," *J. Polit. Econ.*, Feb. 1980, *88*, 174–85.

R. Michener, "An Application of Non-Additive Preferences to a Problem in Money and Growth," work. paper, Univ. Chicago, Nov. 1977.

R. A. Mundell, "Inflation and Real Interest," *J. Polit. Econ.*, June 1963, *71*, 280–83.

J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts," *J. Amer. Statist. Assn.*, June 1960, *55*, 299–306.

T. J. Sargent, "A Classical Macroeconometric Model of the United States," *J. Polit. Econ.*, Apr. 1976, *84*, 207–38.

A. J. Schwartz, "Secular Price Change in Historical Perspective," *J. Money, Credit, Banking*, Feb. 1973, *5*, 243–69.

M. Sidrauski, (1967a) "Rational Choice and Patterns of Growth in a Monetary Economy," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 534–44.

______, (1967b) "Inflation and Economic Growth," *J. Polit. Econ.*, Dec. 1967, *75*, 796–810.

C. A. Sims, "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *J. Amer. Statist. Assn.*, Mar. 1972, *67*, 169–75.

J. Tobin, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–84.

R. C. Vogel, "The Dynamics of Inflation in Latin America, 1950–1969," *Amer. Econ. Rev.*, Mar. 1974, *64*, 102–14.

Peter Whittle, *Prediction and Regulation by Linear Least-Square Methods*, London 1963.

Board of Governors of the Federal Reserve System , *Fed. Res. Bull.*, Washington various issues.

U.S. Bureau of Labor Statistics, *Employment and Earnings*, various issues.

Reswitching, Wicksell Effects, and the Neoclassical Production Function: Note

By KLAUS PERTZ*

In a recent contribution to this *Review*, David Laibman and Edward Nell examined the possibility of reswitching under a variety of assumptions. Using an indecomposable two-sector model with circulating capital, one of these assumptions requires that the principal diagonal of the technique matrix is zero. Laibman and Nell argue that this is indeed a very special but still "...economically meaningful condition strong enough to yield the conclusion that the Clark story requires...." They conclude: "This tells us at best that neoclassical theory can only deal with highly specialized and unusual cases of changes in technique" (p. 882). Although formally correct, if one focuses attention on two-sector models, this line of reasoning needs major qualification.

I

Instead of concentrating on a two-sector model, let us assume a multisector world with one consumption good (produced in sector 1) and with $n-1$ heterogeneous capital goods (produced in sectors $2, 3, \ldots, n$).[1] Then the technique matrix is defined as

$$\left[\begin{array}{c} a_0(\cdot) \\ \hdashline a(\cdot) \end{array}\right] = \left[\begin{array}{cccc} a_{01}(\cdot) & a_{02}(\cdot) & \cdots & a_{0n}(\cdot) \\ \hdashline a_{11}(\cdot) & a_{12}(\cdot) & \cdots & a_{1n}(\cdot) \\ \vdots & \vdots & & \vdots \\ a_{n1}(\cdot) & a_{n2}(\cdot) & \cdots & a_{nn}(\cdot) \end{array}\right]$$

where a_0 represents the vector of direct labor inputs and $(\cdot)=A, B, C, \ldots$ denotes the technique in use.

To examine the possibility of reswitching let us first assume, as in Laibman and Nell, that we have two techniques, A and B, which differ in only one process, say the nth sector. Then we can prove the following theorem:[2]

THEOREM: *Reswitching of process n (the Ruth Cohen Curiosum) is impossible if the difference Δ between both vectors of nonlabor inputs for producing good n is semipositive or seminegative*. (See the Appendix for proof.)

COROLLARY: For reswitching of a single process to exist, there must be at least two different nonlabor inputs in the process.

Now we are able to see the true reason for nonreswitching in the analysis of Laibman and Nell. Since in two-sector models a zero-principal diagonal in the technique matrix happens to coincide with the assumption of only one nonlabor input existing in each sector,[3] Δ is necessarily semipositive or seminegative and, therefore, the Ruth Cohen Curiosum is excluded.

To what extent does this result depend on the assumption that the techniques compared have one process in common? Laibman and Nell argue that it is more realistic to assume that a change in the method of production changes the quality of the product, requiring in turn an adaptation in the activity producing the second good. But a change in the quality of the product necessarily means that the new good must be conceived as a physically different good, the price of which cannot be compared with that of the previously produced one. Otherwise it would follow that the existence of switch points may depend on the choice of numeraire![4]

---

*Johann Wolfgang Goethe-Universität, Frankfurt.

[1] All the other assumptions of Laibman and Nell are retained.

[2] See Joseph Stiglitz, p. 121.

[3] In addition, the requirement "principal diagonal = 0" is unnecessarily strong, since in an indecomposable two-sector model, reswitching is impossible only if the process with the variable technique does not use its own output as input.

II

In order to examine the possibility of reswitching in a two-sector model where both sectors change techniques and no process uses its own output, let us assume an economy comprising three (physically) different goods. System A produces a capital good of a certain type and a consumption good, while system B produces a capital good which is qualitatively different from the one in system A and a consumption good by means of a different method.[5] The input requirements in the process producing the first type of the capital good clearly differ from the coefficients in the process producing the second type of the capital good. For such a model the requirement "principal diagonal = 0" is of no consequence for reswitching as the following example demonstrates:[6]

System A:

$$\left[\begin{array}{c} a_0(A) \\ \hdashline a(A) \end{array}\right] = \left[\begin{array}{ccc} a_{01}(A) & a_{02}(A) & a_{03}(A) \\ \hdashline a_{11}(A) & a_{12}(A) & a_{13}(A) \\ a_{21}(A) & a_{22}(A) & a_{23}(A) \\ a_{31}(A) & a_{32}(A) & a_{33}(A) \end{array}\right]$$

$$= \left[\begin{array}{ccc} 0.8 & 1 & 0 \\ \hdashline 0 & 0.1 & 0 \\ 0.7 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

System B: $$\left[\begin{array}{c} a_0(B) \\ \hdashline a(B) \end{array}\right]$$

$$= \left[\begin{array}{ccc} a_{01}(B) & a_{02}(B) & a_{03}(B) \\ \hdashline a_{11}(B) & a_{12}(B) & a_{13}(B) \\ a_{21}(B) & a_{22}(B) & a_{23}(B) \\ a_{31}(B) & a_{32}(B) & a_{33}(B) \end{array}\right]$$

$$= \left[\begin{array}{ccc} 1.3 & 0 & 0.9 \\ \hdashline 0 & 0 & 0.145 \\ 0 & 0 & 0 \\ 0.5 & 0 & 0 \end{array}\right]$$

Here, system A is used at low and high rates of profit, while for profit rates in the interval $r = 1.17$ to $r = 2.25$, approximately, technique B is selected.[7]

III

I conclude that a zero-principal diagonal is in general irrelevant to reswitching, however defined, not because an economy in which no sector uses its own output as input is "highly specialized" from the empirical point of view, as Laibman and Nell criticize, but because this requirement a priori has nothing to do with reswitching in multisector models with more than two products. In a two-sector world, where one process remains invariant along changes in technique, a zero-principal diagonal has consequences for reswitching, but only because in this very special case it is reduced to the assumption of just one nonlabor input in each sector. Then, clearly, various techniques can be ordered according to their "capital intensity" which can be defined unambiguously as is required by neoclassical doctrine.

---

[4] This can easily be verified by considering the example in Laibman and Nell, p. 883. Using the capital good as numeraire we see that technique A dominates technique B for all feasible values of the rate of profit, although reswitching does occur if one chooses the consumption good as numeraire.

[5] In other words, we have four processes: two processes to produce the same consumption good plus two processes in the capital goods sector, one for each type of the capital good. Accordingly, the number of unknowns is four: the wage rate, the rate of profit, the price of the first capital good, and the price of the second capital good (when, by assumption, the consumption good is used as numeraire). It can be seen that the total number of processes coincides with the total number of unknowns so that the system is determinate.

[6] My notation differs from that of Laibman and Nell which is formally incorrect, since they use the same indices for the input coefficients of physically different goods.

[7] Note that the example could well be handled by augmenting the above matrices in the way suggested by Piero Sraffa, pp. 82–83.

APPENDIX

THEOREM: *Reswitching of process n is impossible if* $\Delta=(\Delta_1,\ldots,\Delta_n)\neq(0,\ldots,0)$ *is semipositive or seminegative, where* $\Delta_i=a_{in}(A)-a_{in}(B)$, $i=1,2,\ldots,n$.

PROOF:

Assume that technique B is optimal for $r=r_1$. Then in steady-state price equilibrium we have $p(B)=[a_0(B)+p(B)a(B)](r_1+1)$, where $p(B)=[p_1(B),\ldots,p_n(B)]$ is the vector of commodity prices with the wage rate as numeraire. The use of technique A in connection with the same price vector $p(B)$ would lead to $p(B)\leqslant[a_0(A)+p(B)a(A)]\cdot(r_1+1)$ which implies $[a_0(A)-a_0(B)]+p(B)[a(A)-a(B)]\geqslant 0$. But if $r=r'$ is a switch point we must clearly have $[a_0(A)-a_0(B)]+p(B)[a(A)-a(B)]=0$ with $p(B)=p(A)$, or in the above notation

$$\Delta_0+p(B)\Delta^T=0 \tag{A1}$$

where $\Delta_0=a_{0n}(A)-a_{0n}(B)$ and T denotes the transpose operator. Obviously, since prices (measured in units of the wage rate) are strictly increasing functions of the rate of profit, there exists at most one $r=r'$ which fulfills (1) if Δ is semipositive or seminegative.

REFERENCES

D. Laibman and E. J. Nell, "Reswitching, Wicksell Effects, and the Neoclassical Production Function," *Amer. Econ. Rev.*, Dec. 1977, *67*, 878–88.

Piero Sraffa, *Production of Commodities by Means of Commodities*, Cambridge 1960.

J. E. Stiglitz, "The Badly-Behaved Economy with the Well-Behaved Production Function," in James A. Mirrlees and Nicholas H. Stern, eds., *Models of Economic Growth*, London; Basingstoke 1973, 117–37.

The Ban on Indexed Bonds, 1933–77

By J. HUSTON MCCULLOCH*

In a recent article in this *Review*, Nissan Liviatan and David Levhari attempt to explain the phenomenon that "In spite of recent marked inflationary trends the important capital markets have not developed a significant free (private enterprise) market for [cost-of-living] linked bonds" (p. 366). They note that the familiar theoretical complete dominance of indexed bonds over nonindexed bonds holds unambiguously only in an economy in which there are no nominal assets. They demonstrate that borrowers may in theory insist on hedging the inflation risk on their nonindexed cash balances by issuing some nonindexed bonds at a discount that makes them attractive in spite of the risk they impose on lenders. Their model leads them to conclude that indexed bonds might not completely dominate nonindexed bonds. However, they do not claim that their model explains the phenomenon they set out to explain: Why is it that the "free market," to use their expression, has come up with virtually no indexed bonds *at all*?

Fortunately, a simple explanation for the nonexistence of indexed bonds in the United States was provided, two years before Liviatan and Levhari's article appeared, by the Supreme Court of the State of Tennessee. In 1974 the Union Planters National Bank of Memphis actually tried to collect on an indexed loan that it had made to one Aztec Properties, Inc., but was thrown out of court for the simple reason that "It would be contrary to the national policy, as expressed by the Congress and as interpreted in several cases by the United States Supreme Court, to permit a lender to require of a borrower a different quantity or number of dollars from that loaned, insofar as the principal amount is concerned" (*Aztec Properties*, p. 760).

The Court based its argument on the so-called "Gold Clause" Joint Congressional Resolution of June 5, 1933:

> Every provision contained in or made with respect to any obligation which purports to give the obligee a right to require payment in gold or a particular kind of coin or currency, or in an amount in money of the United States measured thereby, is declared to be against public policy; and no such provision shall be contained in or made with respect to any obligation hereafter incurred. Every obligation, heretofore or hereafter incurred, whether or not any such provision is contained therein or made with respect thereto, shall be discharged upon payment, dollar for dollar, in any coin or currency which at the time of the payment is legal tender for public and private debts. Any such provision contained in any law authorizing obligations to be issued by or under the authority of the United States, is hereby repealed, but the repeal of any such provisions shall not invalidate any other provision or authority contained in such law.[1]

The first sentence of this law prohibits gold clauses, and does not pertain to indexation. However, most legal scholars have agreed that "the wording of the second sentence of the Resolution is sufficient to invalidate a price-index clause if literally construed" (Gordon McClintock, p. 955), and that this sentence, "read by itself and torn out of its context, would clearly mean... that index clauses would be unenforceable" (Joseph Dach, p. 332).[2]

---

*Associate professor of economics and finance, Ohio State University. I am grateful to G. L. Bach, A. Blinder, J. Buchanan, G. T. Dunne, S. Fischer, M. Friedman, T. M. Humphrey, C. C. Lichtenstein, R. L. Miller, F. Modigliani, R. A. Musgrave, K. S. Rosenn, E. M. Sunley, T. N. Tideman, T. D. Willett, and anonymous referees for helpful comments and suggestions. I remain solely responsible for the contents and conclusions. Publication of this paper was supported by NSF grant SOC 78-13780.

[1] *United States Code Annotated*, Title 31, Section 463.

John Dawson was able to argue that the courts would probably instead interpret the second sentence as an extension of the first, merely specifying the disposition of bonds containing the prohibited gold clauses, so that index clauses were probably still enforceable (p. 683). However, in a 1939 case cited by the Tennessee Court, Hugo Black, speaking for a majority of the U.S. Supreme Court, made it clear that the second sentence was indeed to be "literally construed," "read by itself," and "torn out of context":

> Congress—apparently to obviate any possible misunderstanding as to the breadth of its objective—added, with studied precision, a catchall second sentence sweeping in "every obligation," existing or future, "payable in money of the United States," irrespective of "whether or not any such provision is contained therein or made with respect thereto." The obligations hit at by Congress were those "payable in money of the United States." All such obligations were declared dischargeable "upon payment, dollar for dollar, in any coin or currency [of the United States] which at the time of the payment is legal tender for public and private debts."[3]

In this case, the Court was striking down a foreign currency clause, rather than an index clause. However, the implications for Consumer Price Index (*CPI*) indexation are obvious.

Since 1939 there has been no shortage of legal scholars who valiantly argue in favor of indexation, but they all rely on Dawson's obsolete contention that the second sentence can be read as limited to the obligations enumerated in the first.[4] None of these lawyers directly confronts Justice Black's pronouncement cited above. Nussbaum does mention that in the 1939 case "the statute has been expanded to the breaking point" (p. 307), but the mere outrageousness of a decision does not make it any less binding. Conceivably a later court would have reversed this precedent. However, given the absence of such a case, we must agree with Gerald Dunne that, at least as of his writing,

> ...the half-forgotten Joint Resolution of June 5, 1933 is so fortified by Mr. Justice Black's hard-lining opinion in *Guaranty Trust v. Henwood* that any effort to use indexation as an anti-inflationary device in a loan agreement presents the unwary with all the perils of a rusty bear trap in the high grass. [p. 131]

In short, a sufficient reason for the virtually complete absence of indexed bonds in the United States as of Levhari and Liviatan's writing, and one which was probably much more effective than the interaction with nominally denominated cash balances, is that *CPI* indexation clauses in loan contracts were illegal.[5] If the Joint Resolution had worked by imposing conspicuous criminal sanctions against indexation, a few of the countless economists who have advocated indexation (see Thomas Humphrey) might have been aware of its effect. However, the ban has worked silently yet very effectively simply by making index clauses unenforceable in court. Therefore only a handful of legal specialists need have known, or even have suspected, the ban's existence, in order to have made it effective.

Prior to 1933, gold clause bonds apparently met the public's demand for bonds with a safe purchasing power. Gold was not stable enough in value to please certain economists, but when Irving Fisher's Rand Kardex Corporation actually tried to float an indexed bond in 1925, the public was unenthusiastic (see Dawson and James Coultrap). Gold clauses were not perfect, but they would at least have protected the lender fairly well against a Civil War-type depreciation of the paper dollar. Between 1933 and 1977, both gold clauses and index clauses were unenforceable, and therefore were not used.

---

[2] Note that the third sentence prohibits even the Treasury from issuing the proscribed instruments.

[3] *Guarantee Trust Company, Trustee v. Henwood, Trustee et al.*, 307 U.S. 247, 252, quoted in *Aztec Properties, Inc. v. United Planters*, pp. 760–61.

[4] See Dach (p. 335); Arthur Nussbaum (p. 307); Ellyahu Hirschberg (p. 160–61, 165); Keith Rosenn (p. 741); Hyer and Kearl (pp. 228–31).

[5] The 1975 *Aztec Properties* decision mentioned that other kinds of contracts, such as those governing wages and salaries or rentals, were not affected by the Joint Resolution, and could legitimately be *CPI* indexed (p. 761).

In October 1977, Congress, thinking it was merely humoring a handful of "gold bugs," passed the Helms Amendment, which revoked the 1933 Joint Resolution, "dollar for dollar" wording and all. In doing so, it unwittingly made *CPI* indexation legal. Today, the gold standard is all but forgotten, while price indices have become increasingly refined, available, and familiar to the public. Therefore, indexed bonds probably dominate gold clauses almost completely in today's economy. Unless there are remaining legal obstacles to their use, indexed bonds should soon appear.

State usury laws may be such an additional obstacle. Indeed, the *Aztec Properties* case was somewhat complicated by the fact that the agreed-upon coupons, plus the inflated principal, would have exceeded the Tennessee 10 percent usury law. The Court actually decided that the index clause was doubly unenforceable. McClintock (pp. 973–75), however, argues that as long as the extra interest is subject to a valid contingency (such as the nonappearance of inflation), and especially if it can be construed as compensating for a hazard (such as deflation), usury laws should have no effect.

A more serious potential problem is negotiability. Under Article 8 of the Uniform Commercial Code, indexed bonds would be automatically negotiable as soon as they became commonly traded on the securities exchanges. However, "the anomaly is that until the instruments are in common usage they will not be negotiable, but until they are negotiable they are a poor investment" (McClintock, p. 971). Unless the public decides that the advantage of indexation outweighs the disadvantage of nonnegotiability, indexation may have to wait for a statutory amendment to the existing definition of negotiability. Israel passed such a law in 1957, with the result that "CLI index linked bonds have practically driven unlinked bonds out of the stock exchange" (Hirschberg, p. 163). Indexed bonds also flourished in France prior to the disruptions of World War II (see Rosenn, pp. 745–64), and again after the war until their prohibition under Charles De Gaulle (see Milton Friedman, p. 80).

I might add that a sufficient reason for the continued absence of *CPI* indexed mortgages in the United States since the Helms Amendment, at least as far as federally chartered savings and loan associations go, has been the Federal Home Loan Bank Board's regulations prohibiting them.[6] Michael Hyer and James Kearl discuss this and other legal impediments to mortgage indexation.

In conclusion, the real lesson to be drawn from Liviatan and Levhari's paper is not how invincible is the strength of the Pigou effect. Rather it is how often the apparent irrationality of the free market is actually due to some forgotten "rusty bear trap," to use Dunne's metaphor, hidden in the endless "high grass" of government intervention. The "free" market has not been anywhere near as free as economists find it convenient to assume.

---

[6] Despite the claims of the Board's former Chairman Robert McKinney, these regulations are not justified by the Board's *Alternative Mortgage Instruments Research Study*. Indeed, one contribution to that study actually notes that price level adjusted mortgages were not to be even considered in the study (see Alfred Field and Henry Cassidy).

REFERENCES

J. Dach, "Validity of Price-Index Clauses Under the Gold Coin Joint Resolution," *George Washington Law Rev.*, Apr. 1945, *13*, 328–44.

J. P. Dawson, "The Gold Clause Decisions," *Michigan Law Rev.*, Mar. 1935, *33*, 647–84.

______ **and J. W. Coultrap,** "Contracting by Reference to Price Indices," *Michigan Law Rev.*, Mar. 1935, *33*, 685–705.

G. T. Dunne, "Gold Clause Prohibitions—Alive and Well," *Banking Law J.*, Feb.

1977, *93*, 131–32.

A. E. Field and H. J. Cassidy, "Simulation Analysis of Alternative Mortgage Instruments," in *Alternative Mortgage Instruments Research Study*, Vol. 2, Washington 1977, ch. 8.

M. Friedman, "Gold, Money and the Law: Comment," in Henry G. Manne and Roger L. Miller, eds., *Gold, Money and the Law*, Chicago 1975, 71–81.

E. Hirschberg, "Index Value Clauses—The Israeli Experience," *Banking Law J.*, Feb. 1975, *92*, 158–65.

T. M. Humphrey, "The Concept of Indexation in the History of Economic Thought," *Econ. Rev. Fed. Reserve Bank Richmond*, Nov./Dec. 1974, *60*, 3–16.

M. Hyer and J. R. Kearl, "Legal Impediments to Mortgage Innovation," *Real Estate Law J.*, Nov. 1978, *6*, 211–38.

N. Liviatan and D. Levhari, "Risk and the Theory of Indexed Bonds," *Amer. Econ. Rev.*, June 1977, *67*, 366–75.

G. E. McClintock, "The Probable Legal Consequences of Inserting Price-Index Clauses in Long-Term Corporate Obligations," *Hastings Law J.*, May 1967, *18*, 959–80.

Arthur Nussbaum, *Money in the Law: National and International*, Brooklyn 1950.

K. S. Rosenn, "Protecting Contracts from Inflation," *Business Lawyer*, Jan. 1978, *33*, 729–47.

Aztec Properties, Inc., Appellant v. Union Planters National Bank of Memphis, Appellee, South Western Reporter, 1976, *530 2d*, 756–61.

United States Code Annotated, Title 31: *Money and Finance*, St. Paul 1976.

Inflation Uncertainty and the Phillips Curve: Some Empirical Evidence

By Maurice D. Levi and John H. Makin*

The coincidence of high rates of unemployment and high rates of inflation, seemingly apparent in several countries for sustained periods of time, has led to a rethinking of the Phillips curve. One explanation for the appearance of a positive association between unemployment and inflation, advanced by Milton Friedman in his Nobel lecture, involves an observed linkage between the level and the variability in inflation. This note is an attempt to test whether inflation uncertainty, which should be related to variability, could explain the observed patterns of inflation and employment. It is shown that inflation uncertainty is negatively related to employment, and that failure to consider uncertainty of inflation can obscure the effect of inflation forecast errors on employment.

I. Forecast Errors, Uncertainty, and Employment

The suggestion, originally advanced by Friedman (1968) and by Edmund Phelps, that the Phillips curve is shifted upward by the full amount of anticipated inflation, can be captured—although somewhat crudely—by writing the Phillips curve as

$$(1) \qquad dN_t = b_0 + b_1(dp_t - dp_t^*) + \mu_t$$

where dN_t is the percentage change in employment between the end of $t-1$ and the end of t, dp_t is actual inflation, and dp_t^* is expected or anticipated inflation during t.

We have written (1) in terms of employment rather than *un*employment. This is because the reasoning behind the Phillips curve suggests that an underestimation of inflation should result in more people working. This translates into lower unemployment only in conjunction with the labor force. However, since various factors affect the size of the labor force, including anything that influences participation rates, putting (1) in terms of unemployment creates additional errors. These additional errors, coming from omitting influences on the size of the labor force, do not affect our major conclusions. They do, however, result in serially correlated errors, a phenomenon not uncommonly the result of excluded variables, thereby making interpretation of the results more difficult. We therefore write inflation forecast errors as affecting what Phillips curve arguments suggest—namely, the number employed.[1]

The modified Phillips effect is represented by a positive b_1, and b_0 represents the "normal" growth rate of employment. There would be only a short-run tradeoff since, in a sustained inflation, expectations would become correct (i.e., $dp_t^* = dp_t$) and we would return to the normal rate.

In his Nobel lecture, Friedman has taken the Phillips curve debate into a "third stage" by arguing:

(i) that higher rates of inflation have been associated with higher variability of the rate of inflation;[2] and

---

*Associate professor, Faculty of Commerce, University of British Columbia, and professor, department of economics, University of Washington, respectively. We owe thanks for valuable comments and suggestions to Milton Friedman and for the extremely careful comments of an anonymous referee of this *Review*. We also wish to thank the Fraser Institute. All views expressed and any remaining errors are ours alone.

[1] Analyses of Phillips curves are also frequently expressed in terms of wage-determination relationships (see Robert Gordon, 1976, William Riddell, Stephen Turnovsky and Michael Wachter, and Robert Hall) or income-determination equations (see Robert Lucas, 1972, 1976).

[2] Friedman cites the work of Dennis Logue and Thomas Willett and Dwight Jaffee and Ephraim Kleiman.

(ii) that due to institutional rigidities, higher variability in the rate of inflation has produced a reduction in the efficiency of the price system in guiding economic activity and a consequent reduction in output.

From these two components of Friedman's argument, we have that high inflation, accompanied by high variability of inflation, will mean a low output. The low output could then mean high or low unemployment. If the effect is high unemployment, we have an upward-sloping Phillips curve that relates the level of inflation to unemployment.

We might wish to replace the two arguments used by Friedman by two arguments that are tied more closely to the reformed Phillips curve as given in (1). We could suggest:

(i) that higher errors in predicting inflation $(dp_t - dp_t^*)$ should be associated with higher uncertainty about the rate of inflation, and

(ii) that higher uncertainty about inflation could be associated with lower employment.

A comparison between these two arguments and those of Friedman shows them to be similar although different. Variability and uncertainty about inflation are not the same thing, although in practice they are likely to be related. Similarly, low output will not necessarily be associated with low employment, although, again, they should generally be related.

The arguments presented here, however, are not without intuitive basis and are supported by the data. That more uncertainty about inflation should be associated with larger errors (which have generally been underestimates over the relevant period) should be intuitively acceptable. That higher uncertainty of inflation means lower employment could be defended on Friedman's grounds, namely the lower economic efficiency that accompanies more variable and therefore presumably more uncertain inflation. If the lower efficiency means lower employment, which Friedman says might or might not occur, we have a negative association between uncertainty and employment. We might alternatively argue for a more direct negative link between inflation uncertainty and employment. If higher inflation uncertainty leads to a postponement or cancellation of spending plans, we might directly obtain a negative association.

Allowing for an effect of uncertainty about inflation on employment means modifying our "Phillips curve" in (1) and writing

$$(2) \qquad dN_t = b_0 + b_1'(dp_t - dp_t^*) + b_2\sigma_t + \mu_t$$

where σ_t is a measure of inflation uncertainty that we describe below.

It is not difficult to show that if the two right-hand variables in (2) are positively covariable and if the sign of b_2 is negative—the two components of our argument above—then if we omit σ_t and estimate the modified Phillips curve in its simple form (1), we will be biasing the estimate of b_1 toward zero.

The estimated value of b_1 obtained by the use of (1) would be

$$(3) \qquad \hat{b}_1 = \sigma_{dNE}/\sigma_E^2$$

and the value of $\hat{b}_1'$ from (2) can be written as

$$(4) \qquad \hat{b}_1' = \frac{\sigma_{dNE}}{\sigma_E^2} - b^2\frac{\sigma_{E\sigma}}{\sigma_E^2} = \hat{b}_1 - b_2\frac{\sigma_{E\sigma}}{\sigma_E^2}$$

where we write $E \equiv (dp_t - dp_t^*)$, and σ_{ij} as the covariance between i and j.

Our two arguments for the consideration of uncertainty involve the expectations that $\sigma_{E\sigma} > 0$ (from larger *underestimates* associated with higher uncertainty) and that $b_2 < 0$. If these expectations are true, then it follows from (4) that the positive $\hat{b}_1$ will be larger in magnitude from the estimate via (2) than from the incompletely specified (1), i.e., $\hat{b}_1' > \hat{b}_1$. The measured impact of the degree of error in predicting inflation upon the rate of employment will be larger if account is taken of uncertainty of inflation.

This outcome is not difficult to explain. When actual inflation exceeds anticipated inflation, we expect an increase in employment for the reasons so clearly laid out by Friedman (1968) and Phelps. If the excess of actual inflation is matched by high uncertainty of inflation, this will offset the increase in employment by itself lowering employment. If we do not separately account for the effect of uncertainty, the overall effect will be attributed to the error-in-inflation term, the coefficient of which will consequently be biased downward.

II. Data Employed

To test the arguments we have advanced, use is made of the survey price expectations data collected by Joseph Livingston and reworked by John Carlson. Livingston's data were collected in April and October and were published in June and December in the *Philadelphia Inquirer*. If there was a sharp movement of prices between collection and publication, Livingston adjusted his averages to maintain a "fair" representation of expectations. Carlson reworked Livingston's data to compute expectations in the month of the survey. We use these revised values for both six- and twelve-month forecasts.

In reworking Livingston's original data, Carlson computed the standard deviation across the inflation forecasts of the respondents. At times when there was low uncertainty about inflation, the consensus among respondents should be reflected in a low standard error, and conversely. We can consequently employ these estimates as a convenient direct measure of the degree of uncertainty about inflation.[3] It is our σ.

Employment is measured as the percent change in the seasonally adjusted number of employees on payrolls of nonagricultural establishments between the end of $t-1$ and the end of t. The standard deviation of respondents is measured at the beginning of t, and actual and expected inflation are for the six- or twelve-month periods starting at the beginning of t.

III. Results

The results from estimating both (1) and (2) over two different periods are presented in Table 1. Note that during both the 1965–75 period and the longer 1948–75 period, and over both the six- and twelve-month forecasts, the degree of uncertainty about inflation possesses the expected negative sign and is significant. It is clear that uncertainty about the expected rate of inflation, as measured by the dispersion among Livingston's respondents, has a negative effect upon the rate of employment growth. The value of b_2 in (2) is indeed negative, and the *t*-statistics are very high. See, also, that allowance for uncertainty over both periods raises the measured magnitude of the effect of errors in inflation forecasts given by $\hat{b}_1$. Further, in none of the regressions, when uncertainty is excluded, is the inflation forecast error statistically significant.[4] However, when σ is added, the value of b_1 is significant in every regression other than that with six-month forecasts over the longer period. It appears that the failure to allow for uncertainty, as is done in a form of the Phillips curve like (1), results in a specification bias toward zero. Allowing for uncertainty gives us a better short-run modified Phillips curve tradeoff. This tradeoff, of course, is short run, because if forecasts are rational, over the long run we expect $dp^* = dp$.

To judge the relative importance of the error in inflation forecasts and the degree of uncertainty, β-statistics were computed for both variables.[5] These show in every regression that the uncertainty variable is relatively more important than inflation errors. Table 1 also reveals that the inclusion of the uncertainty variable results in a sizable increase in the adjusted R^2, especially in the 1965–75 subperiod. Without σ, the R^2 remains very close to zero. We also see that adding σ raises the *D.W.* statistic, even though this does not allow us to accept with certainty the null hypothesis of no first-order serial correlation in the 1948–75 results, and for those for 1965–75 without the uncertainty term. In the 1965–75 results obtained with uncertainty, the *D.W.* statistics are in

---

[3] Other examples of the use of this measure can be found in Menachem Brenner, our paper, and Riddell.

[4] This result differs from those obtained by some other researchers, such as Otto Eckstein and Roger Brinner, and Gordon (1972).

[5] For an account of β-statistics, see Arthur Goldberger.

TABLE 1—ESTIMATES OF THE PHILLIPS CURVE

	1965–75		1948–75	
	With σ	Without σ	With σ	Without σ
Six-Month Periods				
$\hat{b}_0$	3.23	1.14	1.54	0.95
	(6.86)	(2.82)	(4.24)	(3.35)
$\hat{b}_1$	0.26	0.08	0.14	0.05
	(2.37)	(0.50)	(1.26)	(0.45)
β_1	0.36		0.17	
$\hat{b}_2$	−1.91		−0.48	
	(−5.35)		(−2.45)	
β_2	−0.81		−0.34	
R^2	0.61	0.01	0.11	0.00
$D.W.$	1.83	1.11	1.49	1.43
Twelve-Month Periods				
$\hat{b}_0$	3.48	1.05	1.42	0.69
	(6.85)	(2.85)	(3.93)	(2.47)
$\hat{b}_1$	0.26	0.13	0.30	0.20
	(2.68)	(0.93)	(2.79)	(1.86)
β_1	0.38		0.36	
$\hat{b}_2$	−2.25		−0.63	
	(−5.43)		(−2.93)	
β_2	−0.78		−0.38	
R^2	0.61	0.04	0.19	0.06
$D.W.$	2.23	1.19	1.57	1.43

Note: Numbers in parentheses are t-statistics; *D.W.* is the Durbin-Watson statistic.

the definite region, and the t-values are very good. Our significant results therefore do not result from serial correlation.

To test whether significance of uncertainty in those results with indefinite *D. W.* statistics is the result of serial correlation, all regressions were reestimated with Cochrane-Orcutt procedure. These results are in Table 2. All regressions were repeated for completeness, even though the 1965–75 results that allowed for uncertainty already had good *D.W.* values. The results are very much the same as those in Table 1, with all *D.W.* statistics in the certain region. There is a significant negative effect of uncertainty on employment. The measured effect of inflation errors when possible serial correlation is removed does not become larger from the inclusion of uncertainty. This differs from the results shown in Table 1. Since serial correlation should not bias coefficient estimates, but rather their estimated variance, this outcome would not be predicted on a priori grounds. This could explain why some earlier researchers like Eckstein-Brinner and Gordon have found effects of inflation forecast errors without allowing for uncertainty. We do notice, however, that the importance of inflation uncertainty, as judged by *betas*, is still much greater than inflation forecast errors. We also see that adjusted R^2 is noticeably improved by allowing for uncertainty.

IV. Conclusions

An effect of uncertainty about inflation on the level of employment has been strongly supported. To the extent that uncertainty and variability of inflation are related, our results provide support for Friedman's description (1977) of the Phillips curve in its third stage. Further, our results suggest that some previous studies of the *U.S.* situation

TABLE 2—ESTIMATES OF THE PHILLIPS CURVE CORRECTING FOR SERIAL CORRELATION

	1965–75		1948–75	
	With σ	Without σ	With σ	Without σ
Six-Month Periods				
$\hat{b}_0$	3.18	0.39	2.49	0.99
	(6.32)	(0.51)	(4.08)	(2.86)
$\hat{b}_1$	0.26	0.38	0.07	0.04
	(2.35)	(2.55)	(0.71)	(0.37)
β_1	0.36		0.09	
$\hat{b}_2$	−1.88		−1.12	
	(−5.10)		(−3.02)	
β_2	−0.80		−0.79	
$\hat{\rho}$	−0.02	0.60	0.38	0.28
	(−0.08)	(3.43)	(3.04)	(2.18)
R^2	0.59	0.32	0.21	0.08
$D.W.$	1.82	1.68	1.91	1.79
Twelve-Month Periods				
$\hat{b}_0$	3.61	0.77	1.59	0.65
	(7.92)	(1.35)	(3.09)	(1.86)
$\hat{b}_1$	0.26	0.27	0.27	0.26
	(3.00)	(1.53)	(2.33)	(2.09)
β_1	0.39		0.34	
$\hat{b}_2$	−2.68		−0.74	
	(−6.31)		(−2.31)	
β_2	0.82		0.45	
$\hat{\rho}$	−0.21	0.46	0.22	0.28
	(−0.99)	(2.46)	(1.67)	(2.16)
R^2	0.63	0.22	0.23	0.15
$D.W.$	1.98	1.86	1.85	1.82

Note: See Table 1.

that have failed to consider inflation uncertainty may have missed the existence of the Phillips curve in its second stage—namely, that underestimates of inflation raise employment, and vice versa.

To the extent that inflation uncertainty persists and causes lower employment, our results tend to support the case for a wider use of indexing of nominal contracts, which should reduce the impact of uncertainty felt on the real sector.

REFERENCES

M. Brenner, "Inflation and Rates of Return on Marketable Securities: First Tests," in Arnold W. Sametz and Paul Wachtel, eds., *Understanding Capital Markets*, Lexington; Toronto 1977, 85–94.

J. A. Carlson, "A Study of Price Forecasts," *Annals Econ. Soc. Measure.*, Winter 1977, *6*, 27–56.

O. Eckstein and R. Brinner, "The Inflation Process in the U.S.," study for the Joint Economic Comm., 92d Cong., 2d sess., 1972.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

______, "Nobel Lecture: Inflation and Unemployment," *J. Polit Econ.*, June 1977, *85*, 451–72.

Arthur S. Goldberger, *Econometric Theory*, New York 1964.

R. J. Gordon, "Aspects of a Theory of Involuntary Unemployment: A Comment," in Karl Brunner and Alan Meltzer, eds., *The Phillips Curve and Labor Markets*, Vol. 1, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 98–119.

———, "Wage-Price Controls and the Shifting Phillips Curve," *Brookings Papers*, Washington 1972, *2*, 385–421.

R. E. Hall, "The Phillips Curve and Macroeconomic Policy," in Karl Brunner and Alan Meltzer, eds., *The Phillips Curve and Labor Markets*, Vol. 1, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 127–48.

D. Jaffee and E. Kleiman, "The Welfare Implications of Uneven Inflation," seminar paper no. 50, Inst. Int. Econ. Stud., Univ. Stockholm, Nov. 1975.

M. D. Levi and J. H. Makin, "Fisher, Phillips, Friedman and the Measured Impact of Inflation on Interest," *J. Finance*, Mar. 1979, *34*, 35–52.

Joseph A. Livingston, surveys published twice yearly, *Philadelphia Sunday Bulletin*, 1948–71; *Philadelphia Inquirer*, 1972 on.

D. E. Logue and T. D. Willett, "A Note on the Relation Between the Rate and Variability of Inflation," *Economica*, May 1976, *43*, 151–58.

R. E. Lucas, "Econometric Testing of the Natural Rate Hypothesis," in Otto Eckstein, ed., *The Econometrics of Price Determination*, Board of Governors of the Federal Reserve System, Washington 1972.

———, "Econometric Policy Evaluation: A Critique," in Karl Brunner and Alan Meltzer, eds., *The Phillips Curve and Labor Markets*, Vol. 1, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 19–46.

E. S. Phelps, "Money Wage Dynamics and Labor Market Equilibrium," *J. Polit. Econ.*, July 1968, *76*, Part II, 687–711.

W. C. Ridell, "The Empirical Foundation of the Phillips Curve: Evidence from Canadian Wage Contract Data," *Econometrica*, Jan. 1979, *47*, 1–24.

S. J. Turnovsky and M. L. Wachter, "A Test of the 'Expectations Hypothesis' Using Directly Observed Wage and Price Expectations," *Rev. Econ. Statist.*, Feb. 1972, *54*, 47–54.

Pareto-Desirable Redistribution in Kind: Comment

By EDGAR O. OLSEN*

In a recent paper in this *Review*, Geoffrey Brennan and Cliff Walsh (hereafter, B-W) claim to have demolished the case for redistribution in kind that has been developed in the literature on Pareto optimal redistribution. According to B-W, the rationale for in-kind transfers in this literature is that in the presence of specific commodity utility interdependence they are necessary to achieve an efficient allocation of resources. They correctly point out that this proposition is not generally true, and in so doing, demolish one argument for redistribution in kind. However, as I shall show in the first section of this comment, this is not the only possible argument to make in favor of the practice.

Brennan and Walsh go beyond the demolition of this one rationale to argue that if redistribution is necessary to achieve an efficient allocation of resources, it should be in the form of lump sum grants. In attempting to prove this theorem, they sharply distinguish between two cases. In both cases society contains a recipient concerned only about his own consumption of N goods. That is, he has a utility function of the form:

$$U_R = U_R(X_R^1, \ldots, X_R^N) \tag{1}$$

In B-W's first case, there is a donor concerned about his own consumption of these goods and the recipient's consumption of one good.[1] Specifically,

$$U_D = U_D(X_D^1, \ldots, X_D^N, X_R^1) \tag{2}$$

Brennan and Walsh correctly argue that in this case it *may* be possible to achieve an efficient allocation of resources without transfers. That is, there exist indifference maps, production functions, and initial endowments such that the initial allocation is inefficient but an efficient allocation can be attained without making the recipient better off. In B-W's second case, the donor is also concerned about the recipient's level of well-being as the recipient judges it and hence has a utility function of the form:

$$U_D = U_D(X_D^1, \ldots, X_D^N, X_R^1, U_R) \tag{3}$$

In this case they argue that if both the specific and general utility interdependencies are Pareto relevant, then transfers are necessary for efficient resource allocation and the appropriate form of redistribution is lump sum. The second section of my comment shows that this theorem is false.

Together the two sections demonstrate that there is no difference between the two cases with respect to whether transfers are justified and whether they should be in kind. In both cases transfers may not be necessary to attain an efficient allocation, the only efficient allocation that can be achieved by means of a lump sum grant from the donor to the recipient involves giving all of the donor's income to the recipient, and (with one weak assumption added) in-kind transfers are necessary to attain an efficient allocation *preferred by everyone to the allocation in the absence of government action.*

---

*Associate professor, department of economics, University of Virginia. I am grateful to Allan Feldman and an anonymous referee for comments which were helpful in revising an earlier draft of this paper.

[1] In fact, if only one other person was concerned about each person's consumption and if the allocation of resources in the absence of transfers is inefficient, then there is every reason to expect that they would privately reach an agreement resulting in an efficient allocation. The case for redistribution through government is that many people are concerned about the consumption of the same person. This utility interdependence may result in some private transfers. However, after all such transfers, it is quite likely that many parties to this externality will value increased consumption of some good by the recipient, but not enough to justify additional unilateral giving. Two-person models of utility interdependence are used to simplify the exposition, and the assumption of no private transfers is added so that the conclusions will be the same as in the many-person model.

I. A Rationale for In-Kind Transfers

The purpose of this section is to develop a rationale for in-kind transfers that is implicit in the literature on Pareto optimal redistribution (see, for example, my 1971 paper, p. 223) but has never been adequately spelled out. The rationale is that in the presence of specific commodity utility interdependence such transfers are necessary to achieve an efficient allocation of resources preferred by everyone to the allocation in the absence of government action, provided that no one is willing to sacrifice all of his own consumption for a finite increase in consumption by others.[2] This theorem will be proved for B-W's first case almost entirely by reference to published results; the proof for their second case is analogous.

It is well-known (see, for example, James Rodgers, pp. 270–71) that in B-W's first case the allocation of resources in the absence of government action is inefficient, assuming no satiation and ignoring corner solutions. Therefore, there are infinitely many efficient allocations preferred by both the recipient and the donor to the initial allocation. Ignoring corner solutions, Rodgers (pp. 271–72) has shown that no efficient allocation can be attained using an unrestricted cash grant. The only corner solution relevant to this discussion (see my 1978 paper, pp. 4–10) is the allocation resulting from giving all of the donor's income to the recipient. Since the recipient is assumed to be completely selfish, this allocation is efficient. However, if the donor is not willing to sacrifice all of his own consumption for a finite increase in consumption by the recipient, then he prefers the initial allocation to this allocation. Therefore, under the conditions stated, redistribution in kind is necessary to achieve an efficient allocation of resources preferred by everyone to the allocation in the absence of government action.

Now consider the case in which preferences are defined by equations (1) and (3). Ignoring corner solutions, a necessary condition for efficient resource allocation is

$$(4)\quad U_{D,1}/U_{D,2}-U_{D,N+1}/U_{D,2}=U_{R,1}/U_{R,2}$$

where $U_{i,k}$ is the partial derivative of the ith person's utility function with respect to its kth argument. In the absence of government action,

$$(5)\quad U_{D,1}/U_{D,2}=U_{R,1}/U_{R,2}=P_1/P_2$$

where P_i is the market price of the ith good. Assuming no satiation, $U_{D,N+1}$ is everywhere positive. As a result, equation (4) is not satisfied at the initial equilibrium. Therefore, the allocation in the absence of government action is inefficient and there are infinitely many efficient allocations preferred by both the recipient and the donor to this allocation. Ignoring corner solutions, no efficient allocation can be attained using an unrestricted cash grant because such grants do not affect equation (5). Beyond this point, the proof is identical to the proof in the preceding case leading to the same conclusion. In-kind transfers are necessary to achieve an efficient allocation of resources preferred by everyone to the allocation in the absence of government action provided that no one is willing to sacrifice all of his own consumption for a finite increase in consumption by others.

II. An Impossible Theorem

In the second section of their paper, B-W claim that if society is composed of people with preferences defined by equations (1) and (3), and if both the specific and general utility interdependencies in (3) are Pareto relevant, then transfers are necessary to achieve an efficient allocation of resources and the appropriate form of redistribution is lump sum. The following counterexample shows that the first part of this theorem is false and the second is mysterious.

---

[2] This proviso is necessary to take account of corner solutions. For a diagrammatic exposition that deals with corner solutions and satiation, see my 1978 paper.

Assume that there are two individuals in society, a donor and a recipient. They consume two goods, education E and other goods X. The donor directly consumes only other goods, but he also has a selfish interest in the recipient's education, cares about the recipient's well-being, and considers the recipient to be the best judge of his own well-being. The recipient cares only about his own consumption of education and other goods. Let X_d and X_r be the quantities of other goods directly consumed by the donor and the recipient, respectively, and E_r be the quantity of education directly consumed by the recipient. Suppose that the preferences of the recipient and the donor can be represented by the functions:

$$U_r = X_r^{.8} E_r^{.2} \tag{6}$$

$$U_d = X_d^{.8} E_r^{.075} U_r^{.125} = X_d^{.8} X_r^{.1} E_r^{.1} \tag{7}$$

Assume that education and other goods are produced at constant costs of \$1 per unit, and that the donor has an income of \$200 and the recipient an income of \$100 per time period. Hence, the budget frontier of this society is

$$X_d + X_r + E_r = 300 \tag{8}$$

Suppose that pretransfer incomes and market prices are unaffected by transfers. Finally, assume that there will be no private transfers between the donor and the recipient.

The results of this example can be depicted in the triangle in Figure 1 whose base is equal to the quantity of education that can be purchased with the total income of the two people and whose height is equal to the quantity of other goods that can be bought with this total income.[3] Each point in the triangle is interpreted in the following way. The vertical distance from the point to the hypoteneuse is the donor's consumption of other goods, the vertical distance from the point to the base is the recipient's consumption of other goods, and the horizontal distance from the point to the other side is the recipient's consumption of education. With this interpretation, each point inside or on the boundary of the triangle represents a feasible allocation of resources and every feasible allocation can be represented by a point in the triangle.

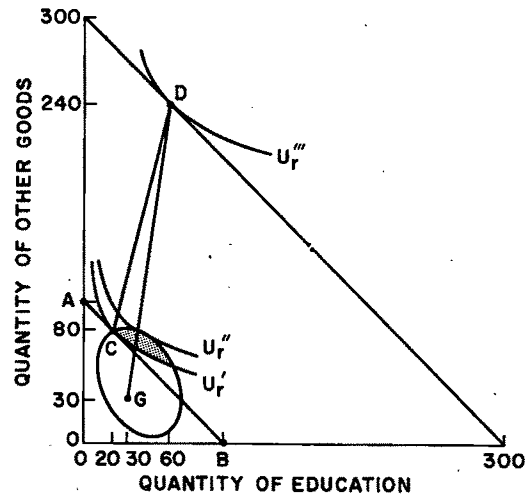


FIGURE 1

The recipient's indifference curves can be depicted in the triangle in the usual way. They are smooth, downward-sloping, convex curves. The donor's indifference curves in the triangle have a circular appearance because they are the projections into (E_r, X_r) space of the intersections in (E_r, X_r, X_d) space of the donor's indifference curves and the societal budget constraint.[4] The point G corresponds to the point where the societal budget plane is tangent to one of the donor's indifference curves.

Line segment AB contains all feasible allocations involving combinations (E_r, X_r) on the recipient's initial budget line. Point C represents the allocation that would prevail in the absence of government action, namely $(E_r = 20, X_r = 80, X_d = 200)$. The recipient would spend 80 percent of his income on

---

[3] This figure is an example of the type of diagram developed by Edgar Browning which in turn is a generalization of the diagrams introduced by Paul Samuelson and F. Trenery Dolbear. Mathematical derivations of all results can be obtained from the author.

[4] The indifference curves are not drawn to be exactly consistent with the assumed utility functions, but they are accurate enough for the present purposes.

other goods and 20 percent on education; the donor would spend all of his income on other goods. The point D is the feasible allocation most preferred by the recipient; the point G the feasible allocation most preferred by the donor. Line segment DC contains all allocations that can be attained by unrestricted cash grants from the donor to the recipient. The line segment DG contains all efficient allocations. Points in the shaded area are preferred by both people to the allocation in the absence of government action.

At the initial allocation C, both the specific and general utility interdependencies are Pareto relevant because a movement to any allocation in the shaded area involves a modification in E_r and U_r that makes the donor better off without the recipient being made worse off. The movement from C to the allocation ($E_r = 36$, $X_r = 72$, $X_d = 192$) is such a movement. Therefore, all of the conditions of B-W's theorem hold, but transfers are not necessary to achieve an efficient allocation of resources. Since the conflict curve DG intersects the recipient's initial indifference curve U_r', it is possible to attain an efficient allocation without making the recipient better off. Since it intersects the recipient's initial budget line, an efficient allocation can be attained without reducing the donor's own consumption.

Figure 1 also illustrates the results of the preceding section. The only efficient allocation of resources attainable using an unrestricted cash grant is that resulting from giving all of the donor's income to the recipient. Clearly, the movement from the initial allocation C to the allocation D is not a Pareto optimal change. These results hold true even if transfers are necessary to attain an efficient allocation. If, for example, the donor's preferences are given by

$$U_d = X_d^{.4} E_r^{.225} U_r^{.375} = X_d^{.4} X_r^{.3} E_r^{.3} \tag{9}$$

instead of equation (7), the feasible allocation most preferred by the donor is represented by a point to the northeast of the recipient's initial indifference curve and hence the conflict curve does not intersect this indifference curve. In this case, it is possible to make both people better off using unrestricted cash grants, but D is the only efficient allocation that can be attained by this means and the donor prefers C to D.

To argue for the exclusive use of lump sum grants in B-W's second case is to argue that the efficient allocation least preferred by the donor should be attained or that an efficient allocation should not be attained. Since I doubt that B-W would accept either of these arguments, I conclude that their second section contains errors in logic.

REFERENCES

G. Brennan and C. Walsh, "Pareto-Desirable Redistribution in Kind: An Impossibility Theorem," *Amer. Econ. Rev.*, Dec. 1977, *67*, 987–90.

E. K. Browning, "The Diagrammatic Analysis of Multiple Consumption Externalities," *Amer. Econ. Rev.*, Sept. 1974, *64*, 707–14.

F. T. Dolbear, "On the Theory of Optimum Externality," *Amer. Econ. Rev.*, Mar. 1967, *57*, 90–103.

E. O. Olsen, "Subsidized Housing in a Competitive Market: Reply," *Amer. Econ. Rev.*, Mar. 1971, *61*, 220–24.

______, "The Simple Analytics of the Externality Argument for Redistribution," unpublished paper, Univ. Virginia, Feb. 1978.

J. D. Rodgers, "Distributional Externalities and the Optimal Form of Income Transfers," *Publ. Finance Quart.*, July 1973, *1*, 266–99.

P. A. Samuelson, "A Diagrammatic Exposition of a Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1955, *37*, 350–56.

Pareto-Desirable Redistribution in Kind: Reply

By GEOFFREY BRENNAN AND CLIFF WALSH*

Edgar Olsen's comments have two purposes: first, to provide an "alternative rationale" for redistribution in kind not susceptible to our criticisms of the orthodox formulation; and second, to disprove our claim that where both "general" and "specific" utility interdependencies are Pareto relevant, transfers are necessary to achieve optimality and need only be lump sum in form. In our view, however, Olsen's comments reveal a failure in the first case to understand our logic, and in the second to define "Pareto relevance" appropriately. In part, the object of our reply is to support these claims, an exercise we address in Section I. The reply also however provides us with the opportunity to correct our earlier statement. It turns out that our original impossibility theorem *is* technically wrong, though for reasons unrelated to Olsen's arguments. The implications of this error are not entirely clear to us. It may be that what we have uncovered is simply an analytic curiosum. On the other hand, it may be that we are able to rehabilitate the concept of Pareto-desirable redistribution in kind, though in circumstances much more restricted than those Olsen and other in-kind transfer exponents have had in mind.

I

Although Olsen accepts our point that, in general, transfers are not necessary to internalize specific commodity utility interdependencies (standard externalities), he claims to offer an alternative rationale for redistribution in kind. In doing so, however, he seems totally to ignore the underlying logic of our argument.

In internalizing an externality associated with R's consumption of good X, "transfers" to R may be *consistent with* the Pareto criterion, but only a substitution effect is strictly *required* by that criterion. Appropriately applied, the phrase "Pareto-desirable redistribution" is, we argue, restricted to situations in which the Pareto criterion *logically requires* an increase in R's utility—in which, in other words, it is logically impossible to soak up all possible gains from trade without increasing R's utility. As we indicate,[1] to argue otherwise would be to accept that *any* situation in which individuals are made better off, *including voluntary trades*, involves redistribution. Such usage would, we believe, merely promote semantic confusion, and it was for this reason that we insisted that the characteristic feature of Pareto-desirable *redistribution* is that the utility-possibilities frontier is upward sloping in the relevant range. Then, and only then, would an increase in R's utility be logically required to reach an efficient point, and it is the minimal necessary increase in R's utility which strictly qualifies as redistribution.[2]

Olsen's alternative rationale is nothing more than a constrained version of the Pareto criterion, such that *increases* in everyone's well-being are *always* required. Thus, he says, "... transfers are necessary to achieve an efficient allocation of resources *preferred by everyone*" (p. 1029, emphasis added). It is true that under *this* criterion income effects are required:[3] indeed they

*Professor of economics, Center for the Study of Public Choice, Virginia Polytechnic Institute and State University, and professor of economics, University of Adelaide, respectively. The paper was prepared while Walsh was senior lecturer in economics, Monash University.

---

[1] The reader's attention is here drawn to our original paper, particularly fn. 8, p. 988.

[2] In an important sense, this redistribution to R is *costless*. Conceptually, it involves making R better off, all other utilities constant, as our later Figure 1 may help to see. Beginning at F, the Paretian redistribution is the increase in R's utility from $U_R^{C'}$ to U_R^G, D's utility constant. This transfer then opens up the prospect of more or less conventional mutual gains, though, in practice, the gains to R may be coincidental with increasing D's utility.

[3] In principle, of course, the required transfer may be infinitesimally small so that even the "Olsen criterion"

would be required in the correction of monopoly problems, the provision of public goods, and presumably in voluntary trades too. However they are required not by the Pareto criterion itself, but rather to satisfy an additional non-Paretian constraint, and *these* income effects have no connection with the essence of the *Pareto-desirable* redistribution literature. If Olsen's alternative rationale is implicit in that literature (as he claims), then it has been entirely misplaced; as ought to have been clear from our original contribution, his rationale is in essence no different from the "orthodox position" we discussed and rejected there.[4] Indeed, even if the required income effects were to be called "Olsen-desirable *redistributions*" (a position we would reject given our previous argument on the use of the word redistribution) we would insist that they have more in common with the conventional *zero sum* redistributions associated with the application of external norms (for example, a completely specified social welfare function) than with the *positive sum* (efficiency-based) redistributions appropriately described as Pareto desirable. For in Olsen's framework, the necessary gains to *R* are, in general, obtained at someone else's expense. In the case of Pareto optimal redistribution as we have defined it, however, transfers to *R* are effectively "costless" to others.[5] Of course, Olsen may dispute our semantics. But it seems to us that our definition of Pareto optimal redistribution is clear, coherent, and unlikely to court confusion: Olsen's alternative usage has none of these features.

We turn now to Olsen's elaborate counterexample to our proposition that if both specific and general utility interdependencies are Pareto relevant, an income effect is required for optimality. His counterexample fails because our conditions are not met: the general utility interdependence is entirely inframarginal over the range in which the individuals are placed, and hence cannot be Pareto relevant. To see this, all we need to ask is whether from *C*, in Olsen's diagram, increases in U_r, *ceteris paribus*, make *D* better off. The answer is no, and Olsen's demonstration depends crucially on this. However, if *G* lies above and to the right of *C*, then the general utility interdependence is no longer inframarginal and the Pareto criterion *does* require that *R* be made better off. Whatever definition of Pareto relevance Olsen is using, it is certainly not the one in common use.

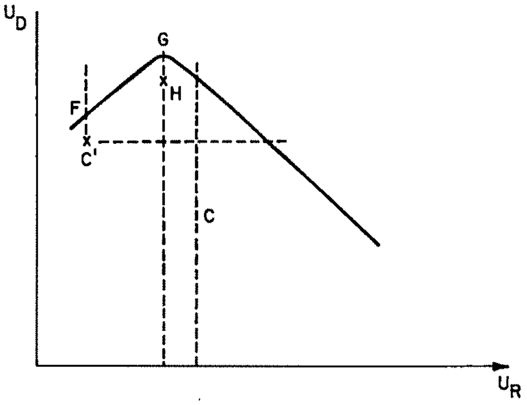


FIGURE 1

If we translate the Olsen example into utility-possibilities space, we obtain something like Figure 1, in which *C* lies to the right of *G*. For our case to apply, we need an initial income distribution that locates us at, say, *C′*, where *both* the conventional externality and the general utility interdependence are Pareto relevant. From *C′* it is clear that *R* must be made better off if optimality is to be achieved:[6] he must at

---

would not *logically require* a major transfer (in-kind) program.

[4] See specifically our original paper pp. 988–89.

[5] See fn. 2 above for a justification of our reference to Paretian transfers as costless. We must acknowledge that Olsen's criterion would admit strict Paretian transfers—but since its purpose is to establish a case for transfers over and above these, this fact has no bearing on the points we make.

[6] One might at this point raise a semantic quibble about our earlier definition. At *C′*, it is not true that making *D* better off *requires* making *R* better off since one can move to *F*. The move to *G* would then, of course, involve Pareto-desirable redistribution (and must, incidentally, be affected via lump sum transfers); but what of the move from *C′* to *H*? To cover this minor ambiguity it would perhaps be preferable to define Pareto-desirable redistribution as a situation in which a move to *optimality* (rather than simply any Pareto-desirable move) requires that *R* be made better off.

least have utility level U_R^G, and the move from $U_R^{C'}$, which is the only redistribution required by the Pareto criterion, *can* and *must* be effected solely in cash.

II

In our earlier paper, and in the foregoing discussion, we asserted that upward-sloping segments of the utility-possibilities frontier only occur with general utility interdependence, and on that basis we rejected the possibility of Pareto-desirable redistribution in kind. We were not quite correct: public goods and externalities *may* involve upward-sloping portions of the utility-possibilities frontier in certain "corner cases" (though not those mentioned by Olsen). The conditions under which this is so are rather stringent and almost certainly *not* what Olsen and other transfer-in-kind exponents have had in mind—but it is important to identify the relevant conditions both in the interests of analytic completeness and because they cannot be held to be questionably policy irrelevant.

A. *The Basic Analysis*

The nature of the relevant corner cases can be most easily isolated for the case of a public good.[7] The simple two-person (A, B), two-commodity (X_1, X_2) world illustrated in Figure 2 suffices for this purpose.

Suppose, initially, that only commodity X_1 (a strictly private good) is produced at its maximum (efficient) output of T units. Individual A, given full command over that output, would attain the utility level represented by i_A^1, while B would have zero utility. Similarly, maximum utility for B is represented by i_B^1, in which case a would have zero utility. Other distributions of the available output T of X_1 would yield the utility combinations depicted in the *strictly downward-sloping* utility-possibilities frontier $U_A^1U_B^1$ in Figure 3.

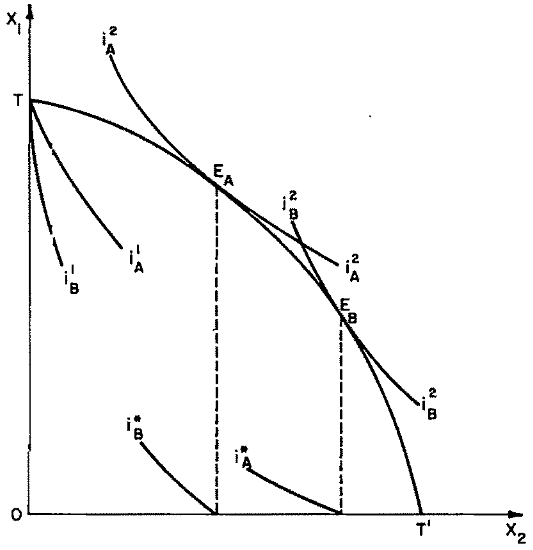


FIGURE 2

Now we introduce the second commodity X_2, supposing initially that it, too, is a private good, and that the production-possibilities frontier is TT' in Figure 2. Individual A's utility maximum is now represented by i_A^2 (with U_B necessarily zero), and B's by i_B^2(with U_A at zero), and the complete utility-possibilities frontier can be depicted in Figure 3 by (the again *strictly downward-sloping*) $U_A^2U_B^2$.

However, if we now suppose that X_2 is not a private good, but rather has the jointness characteristic of a pure public good, the outcome may change in a significant way. Given unchanged production possibilities, A's maximum feasible utility level is still i_A^2. In the private goods case, b necessarily has zero utility because A exhausts the output of both X_1 and X_2. However with X_2 a public good, whatever A consumes of X_2 is fully and equally available for B. Providing that B does not regard units of the public good X_2 as useless in the absence of the private good, his utility is now nonzero (represented by I_B^*) when A's is at its maximum: to reduce B's utility to zero would also involve reducing A's utility to i_A^1 in Figure 2.

---

[7] Strictly speaking, the result with which we are concerned turns only on the jointness characteristic of the good and may in many cases be extended to generalized utility interdependence. In this sense our use of the terms public and private goods is somewhat loose, but convenient.

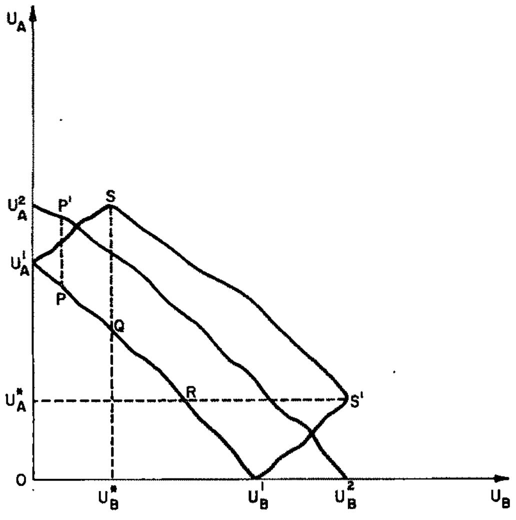


FIGURE 3

By the same token, at B's utility maximum, i_B^2, A's utility is indicated by i_A^* and can only be reduced by also lowering B's utility. This is reflected in Figure 3, in the new utility-possibilities frontier $U_A^1SS'U_B^1$, which over some ranges is *positively sloped*.[8] Only in the downward-sloping range SS' is full Pareto optimality achieved, and the extent of this range depends crucially on the preferences of both A and B for the public good relative to the private good, including the utility level each attains when his private good consumption is zero.

B. *An Interpretation*

Evidently, the provision of any conventional public good may give rise to the positively sloped segments of the utility-possibilities frontier that uniquely generate redistribution that is *required* by the Pareto criterion, and which we have hitherto regarded as arising only for the case of general utility interdependence. In the case illustrated in Figures 2 and 3, the very process of providing the public good may place limits on the distribution of the gains which arise from its availability in a way which is not true of private goods.[9] For any initial distribution on $U_A^1U_B^1$ outside the range QR, efficient provision of the public good, in remaining faithful to the Pareto criterion, specifically *requires* that both of the individuals be made better off. Thus, provision of public goods may be considered to be associated with a form of Pareto-desirable redistribution: and since the redistribution in question cannot be dissociated from the provision of a particular good, it would seem appropriate to refer to this phenomenon as a case of Pareto-desirable redistribution in kind.

It is apparent, furthermore, that analogous results may arise for any form of utility interdependence. Assume, for example, that the "well off" feel that their well-being is adversely affected by the health risks associated with the poor nutritional standards adopted by lower-income groups (i.e., we have an externality of standard sort). Whether the relevant Pareto optimum is achieved by subsidizing consumption of appropriate foods, or by taxing the poor's consumption of other things (plus an exactly utility-compensating lump sum repayment) is, from a Paretian viewpoint, a matter of indifference. If the subsidy method happens to be chosen, it must in general be regarded as a combination of a necessary substitution effect and an *optional* income effect. Consider, however, the alternative option of imposing a tax on the poor's consumption of other things. Then there may remain a Pareto-relevant externality even when the poor are spending *all of their income* on the externality-generating goods. In this case, further gains to the well off could only be secured by additional transfers of the externality generating goods to the poor, potentially making *both* groups better off.

It is conceivable, therefore, that redistribution in kind may be required by the Pareto criterion in association with the internalization of *any* relevant utility interdependence. But the conditions under which this is so are fairly severe. It requires first that individuals can "live by the externality-generating good X alone" (i.e., that additional units of X do add to utility when X is all that is consumed); and second that recipients spend *all* of their income on the externality-generating good. Although these requirements are stringent, the possibility that such redistribution in kind might be policy relevant cannot be totally rejected, particularly if Pareto-relevant externalities apply simultaneously over a significant range of basic consumption goods (health, education, nutritionally desirable foods, etc.). And the possibility ought, at least, be acknowledged for reasons of analytic completeness. If this possibility is viewed as a relevant one, then we have supplied an alternative basis (on a sound analytic footing) for Pareto-desirable redistribution in kind: our original theorem was not only false, but misleading for policy purposes. If the possibility is regarded as remote, however, then the theorem was wrong but not outrageously so: Pareto-desirable redistribution in kind would have to be regarded as not strictly impossible, but simply irrelevent.

---

[8] The jointness characteristic of the public good also leads the utility-possibilities frontier to be further "expanded" than when X_2 is purely private. This fact plays no role in our present analysis, and could be eliminated by assuming that the marginal cost of public good production of X_2 is higher than that of private good production of X_2 by an appropriate offsetting amount.

[9] This fact seems to have escaped Paul Samuelson's attention in his seminal contributions to the public goods literature (see his chart 4, p. 352).

REFERENCES

G. Brennan and C. Walsh, "Pareto-Desirable Redistribution in Kind: An Impossibility Theorem," *Amer. Econ. Rev.*, Dec. 1977, *67*, 987–90.

E. O. Olsen, "Pareto-Desirable Redistribution in Kind: Comment,"*Amer. Econ. Rev.*, Dec. 1980, *70*, 1028–31.

P. A. Samuelson, "A Diagrammatic Exposition of a Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1955, *37*, 350–356.

The Pigouvian Tax Rule under Monopoly

By A. H. BARNETT*

Over the last decade, significant strides have been made in clarifying the issue of taxation for control of external effects. William Baumol and Wallace Oates have shown that, contrary to the position taken by Ronald Coase, James Buchanan and W. Craig Stubblebine, among others, unilateral taxes are appropriate for the control of public externalities. They have further shown that the majority of policy-relevant externalities are of the public variety.[1] Yet these efforts fail to deal adequately with an analytic complication which arises with the Pigouvian approach. In particular, they fail to deal adequately with the problem of taxing externalities when those being taxed are imperfectly competitive firms. Buchanan has described the nature of the problem which arises when externality-generating firms are monopolistic, and Baumol and Oates have restated and extended Buchanan's work. But these analyses are incorrect on some points and incomplete on others. The purpose of this note is to further extend Buchanan's work and to correct an error made by Baumol and Oates.

I. The Model

Two sources of misallocation can occur with imperfectly competitive polluters. One is the distortion due to the externality, and the other is the underproduction of final products generally associated with the exercise of monopoly power. A tax on pollutants will reduce the generation of external damages, but it may also cause firms to reduce further their production of final products. Thus there may be tradeoffs between the two distortions, one due to monopolistic underproduction and the other due to external diseconomies. A tax based only on marginal external damages ignores the social cost of further output contraction by a producer whose output already is below an optimal level.

An ideal solution to this problem would incorporate two policy actions: a device to increase production of final products together with a tax to control the external diseconomy. It is assumed, however, that the product market distortion cannot be directly corrected, and so the pollution tax must achieve an optimal second best tradeoff of distortions.

To see more clearly the nature of this optimal second best tradeoff, let us assume that the externality in question is air pollution. Consider a polluter who produces a single product output q with the (inverse) demand function $f(q)$, and who discharges smoke s generating $e(s, X)$ in external damages: where X is a vector of activity levels x_i for parties damaged by smoke. Assuming that all parties affected adversely by smoke are utility-maximizing individuals who view any tax imposed on pollution as a parameter, and further assuming that all prices relevant to purchasing decisions by these individuals remain unchanged as the rate of smoke production, s, varies, then

$$x_i = g(s) \text{ for all } x_i \text{ in } X$$

Thus, $e(s, X)$ can be written as a function of s alone, i.e.,

$$e(s, X) = E(s)$$

Let total resource costs for the smoking-generating firm be represented by $c(q, w)$:

---

*Assistant professor of economics, University of South Carolina. I am indebted to William Baumol, James Bradley, Michael Connolly, Wallace Oates, Edgar O. Olsen, R. Blaine Roberts, and Roger Sherman for helpful comments.

[1] In general, private externalities will persist only when (i) legal or institutional restrictions prevent bargaining, or (ii) the externality is either insignificant or the cost of price rationing is very high. For a discussion of these points, see Baumol and Oates, pp. 20-21.

where w represents resources devoted to smoke treatment. Assume initially that the firm has two ways of reducing its effluent s. It may either reduce output q, or it may devote more resources w to the treatment of smoke once it is produced.

Now let us consider a tax on effluent. Assuming that all conditions for optimal production are met elsewhere in the economy and that society is indifferent to purely redistributional effects, we can use producer's and consumer's surplus as welfare measures.[2] Social welfare will be given by the difference between the sum of producer's and consumer's surplus and any technological external costs which are not accounted for in producer's surplus. To maximize welfare by taxation, we must find some tax rate per unit of smoke discharged t, which will maximize

$$(1)\qquad u=\int_0^q f(q)dq-c(q,w)-E(s)$$

where q, w, and hence s can be written as functions of t. Differentiating equation (1) with respect to t yields the following first-order condition for welfare maximization:

$$(2)\quad f(q)\frac{dq}{dt}-\frac{\partial c(q,w)}{\partial q}\frac{dq}{dt}-\frac{\partial c(q,w)}{\partial w}\frac{dw}{dt}$$
$$-\frac{dE(s)}{ds}\left[\frac{\partial s}{\partial q}\frac{dq}{dt}+\frac{\partial s}{\partial w}\frac{dw}{dt}\right]=0$$

Profits for the firm are given by

$$(3)\qquad \Pi=f(q)q-c(q,w)-s\cdot t$$

Assuming the polluter views t as a parameter, first-order conditions for profit maximization are

$$(4)$$
$$\frac{\partial\Pi}{\partial q}=f(q)+\frac{df(q)}{dq}q-\frac{\partial c(q,w)}{\partial q}-\frac{\partial s}{\partial q}t=0$$

and

$$(5)\qquad \frac{\partial\Pi}{\partial w}=0-\frac{\partial c(q,w)}{\partial w}-\frac{\partial s}{\partial w}t=0$$

Rewriting (4) and (5), we have

$$(6)\qquad \frac{\partial c(q,w)}{\partial q}=f(q)+\frac{df(q)}{dq}q-\frac{\partial s}{\partial q}t$$

$$(7)\qquad \frac{\partial c(q,w)}{\partial w}=-\frac{\partial s}{\partial w}t$$

Substituting (6) and (7) into equation (2) and simplifying yields a first-order condition for welfare maximization combined with profit-maximizing behavior:

$$(8)\quad 0=-\frac{df(q)}{dq}\frac{dq}{dt}q+\frac{\partial s}{\partial q}\frac{dq}{dt}t+\frac{\partial s}{\partial w}\frac{dw}{dt}t$$
$$-\frac{dE(s)}{ds}\left[\frac{\partial s}{\partial q}\frac{dq}{dt}+\frac{\partial s}{\partial w}\frac{dw}{dt}\right]$$

and the welfare-maximizing tax is

$$(9)\qquad t^*=\frac{\dfrac{df(q)}{dq}\dfrac{dq}{dt}q}{\dfrac{\partial s}{\partial q}\dfrac{dq}{dt}+\dfrac{\partial s}{\partial w}\dfrac{dw}{dt}}+\frac{dE(s)}{ds}$$

Obviously, (9) is not an explicit solution for t because t is on both sides of the equation. But (6) and (7) allow us to write q and w as functions of t. Substituting into (2) then gives one equation in one unknown, t.

II. An Illustration of Two Special Cases

It is now possible to illustrate with (9) two special cases found in the literature concerned with taxation for pollution control. The first case is one in which the only means of smoke abatement is reducing output q and the second case is one in which waste treatment is the only means of smoke abatement. The first of these special cases is often analyzed in the theoretical literature concerned with taxation.[3] The second has been used in several empirical studies.[4] If it is assumed that the only way for a polluter to reduce emissions is by reducing output, then terms involving w in (9) disappear and an optimal tax is given by

---

[2] To be strictly correct, I should use a real income constant measure of consumer's surplus, such as equivalent or compensating variation. The area under a money-income constant demand curve is only approximately correct as a welfare measure.

[3] See, for example, Baumol, Baumol and Oates, Earl Thompson and Ronald Batchelder, and James Marchand and Keith Russell.

$$t^* = \frac{\dfrac{df(q)}{dq}\dfrac{dq}{dt}q}{\dfrac{\partial s}{\partial q}\dfrac{dq}{dt}} + \frac{dE(s)}{ds} \tag{10}$$

On the other hand, if we assume that waste treatment is the only means of smoke abatement, then dq/dt is zero and an optimal tax is given by

$$t^* = \frac{\dfrac{dE(s)}{ds}\left[\dfrac{\partial s}{\partial w}\dfrac{dw}{dt}\right]}{\dfrac{\partial s}{\partial w}\dfrac{dw}{dt}} = \frac{dE(s)}{ds} \tag{11}$$

The term $(df(q)/dq)(dq/dt)$ reflects in part the market structure within which the polluter operates, and, therefore, the significance of the welfare cost associated with reducing the firm's output by imposing emission taxes. If it is assumed that the polluter responds to a tax on emissions by reducing q only, then dq/dt is negative and demand elasticity is relevant for determining an optimal second best tax. Of course, if the polluter responds to a tax by changing w only, then dq/dt is zero and market structure is not a relevant consideration. I shall now interpret condition (9) for the more general case where both q and w vary with t.

III. A More General Case

The importance of market structure in setting optimal pollution taxes may be seen more clearly by rewriting equation (9) to show explicitly the role of price elasticity of demand. The price elasticity of demand for q, η, is

$$\eta = \frac{dq}{df(q)}\frac{f(q)}{q}$$

thus $\dfrac{df(q)}{dq}q = \dfrac{f(q)}{\eta}$

and, of course, $-\dfrac{df(q)}{dq}q = \dfrac{f(q)}{|\eta|}$

Substituting this identity into equation (9) yields the following expression for an optimal second best tax:[5]

$$t^* = \frac{-\dfrac{f(q)}{|\eta|}\dfrac{dq}{dt}}{\dfrac{\partial s}{\partial q}\dfrac{dq}{dt} + \dfrac{\partial s}{\partial w}\dfrac{dw}{dt}} + \frac{dE(s)}{ds} \tag{12}$$

Assuming that $dq/dt < 0$ and $dw/dt > 0$,[6] and

---

[4] See R. Carbone and J. R. Sweigart (1976) and Carbone et al. (1978).

[5] The relevant measure of elasticity is that associated with the post-tax levels of the terms in equation (12).

[6] While these assumptions are common in the literature, they may not be universally applicable. To illustrate, consider the following expressions for dq/dt and dw/dt derived from the standard second-order conditions for profit maximization:

(a) $$\frac{dq}{dt} = \left\{\frac{\partial s}{\partial q}\left[-\frac{\partial^2 c(q,w)}{\partial w^2} - \frac{\partial^2 s}{\partial w^2}t\right] - \frac{\partial s}{\partial w}\left[-\frac{\partial^2 c(q,w)}{\partial q \partial w} - \frac{\partial s}{\partial q \partial w}t\right]\right\} + |D|$$

(b) $$\frac{dw}{dt} = \left\{-\frac{\partial s}{\partial q}\left[-\frac{\partial^2 c(q,w)}{\partial q \partial w} - \frac{\partial^2 s}{\partial q \partial w}t\right] + \frac{\partial s}{\partial w}\left[\frac{d^2 R}{dq^2} - \frac{\partial^2 c(q,w)}{\partial q^2} - \frac{\partial^2 s}{\partial q^2}t\right]\right\} + |D|$$

where R is total revenue and D is the conventional matrix of partial derivatives formed from second-order conditions for profit maximization. Satisfying second-order conditions requires $|D| > 0$ and

$$\frac{d^2 R}{dq^2} - \frac{\partial^2 c(q,w)}{\partial q^2} - \frac{\partial^2 s}{\partial q^2}t < 0$$

Further, $\partial s/\partial q > 0$ and, if we assume diminishing marginal productivity for resources devoted to waste treatment, $\partial s/\partial w$, $\partial^2 c(q,w)/\partial w^2$, and $\partial^2 s/\partial w^2$ are all negative. However, we have no a priori information concerning the signs of $\partial^2 c(q,w)/\partial q \partial w$ and $\partial^2 s/\partial q \partial w$. If the sum of these two terms is negative, or if positive but small relative to the other terms in (a) and (b), then $dq/dt < 0$ and $dw/dt > 0$, as assumed. But, if the sum of $\partial^2 s/\partial q \partial w$ and $\partial^2 c(q,w)/\partial q \partial w$ is positive and large, it is possible that $dq/dt > 0$ or $dw/dt < 0$. Hence, it is possible, though the chance seems remote, that the optimal tax rate would be higher for monopolistic than for competitive firms.

observing that dq/dt approaches some finite number as $|\eta|$ approaches infinity, it is easily seen that as demand approaches the perfectly elastic state the value of the optimal tax rate approaches marginal external damages, $dE(s)/ds$. But when η is finite, a tax rate less than marginal external damages is required to achieve an optimal tradeoff between the external diseconomy and the welfare loss associated with monopoly restricted output. Thus, the optimal second best tax rate can be higher as demand is more price elastic.[7] Moreover, a positive tax will be appropriate only when marginal external damage, $dE(s)/ds$, exceeds

$$\frac{\dfrac{f(q)}{|\eta|}\dfrac{dq}{dt}}{\dfrac{\partial s}{\partial q}\dfrac{dq}{dt}+\dfrac{\partial s}{\partial w}\dfrac{dw}{dt}}$$

The conclusion that optimal second best tax rates can be higher for firms with more price elastic demand functions differs from the conclusion arrived at by Baumol and Oates. In expanding on Buchanan's initial discussion of monopoly and pollution taxes, they state:

> The more inelastic demand, the smaller, apparently, is the change in output resulting from a given change in the [emission] tax. This seems to suggest that other things being equal, we should adopt a higher emission tax where the product demand is *more price inelastic*. As is widely recognized, this role of elasticity is typical of many second-best problems of pricing and resource utilization.
> [p. 77, emphasis added]

Baumol and Oates fail to note here a difference between most second best problems and the special problem of taxing for pollution control. In the usual case, social losses from taxation are minimized by minimizing distortions in market-dictated resource utilization. However, the purpose of a tax on pollutants is to reduce the social cost of market imperfections where the starting point is not ideal. Corrective taxes reduce external costs but they add to the deadweight loss attributable to the monopolist's restricted output.

IV. Conclusions

In the search for efficient controls for externalities, economists have, in general, supported taxes over administrative regulations. The reasons for this choice are well known and grounded solidly in economic theory. However, in our enthusiasm for market-like solutions, we have been remiss in failing to sort out the complexities of optimal taxation. This paper has dealt with one factor which complicates the derivation of optimal taxes for the control of externalities. I have derived two main conclusions: (i) where polluters are imperfectly competitive, second best optimal tax rates may be less than marginal effluent harm; and (ii) the amount by which optimal tax rates fall short of marginal damages may increase as price elasticity of demand for the polluter's produce decreases.

REFERENCES

William J. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

______ **and Wallace E. Oates,** *The Theory of Environmental Policy*, Englewood Cliffs 1975.

J. M. Buchanan, "External Diseconomies, Corrective Taxes, and Market Structures," *Amer. Econ. Rev.*, Mar. 1969, *59*, 174–77.

______ **and W. C. Stubblebine,** "Externality," *Economica*, Nov. 1962, *29*, 371–84.

R. Carbone et al., "A Bargaining Resolution of the Efficiency versus Equity Conflict in Energy and Air Pollution Regulation," in J. S. Arnofsky and A. G. Rao, eds., *Energy Policy*, Amsterdam 1978.

______ **and J. R. Sweigart,** "Equity and Selective Pollution Abatement Procedures,"

---

[7] It should be noted, however, that the optimal tax rate need not increase monotonically with price elasticity of demand.

Manage. Sci., Dec. 1976, *23*, 361–80.

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

C. J. Goetz and J. M. Buchanan, "External Diseconomies in Competitive Supply," *Amer. Econ. Rev.*, Dec. 1971, *61*, 883–90.

J. R. Marchand and K. P. Russell, "Externalities, Liability, Separability, and Resource Allocation," *Amer. Econ. Rev.*, Sept. 1973, *63*, 611–20.

E. A. Thompson and R. Batchelder, "On Taxation and the Control of Externalities: Comment," *Amer. Econ. Rev.*, June 1974, *64*, 467–71.

A Note on the Separability of Production and Location

By ARTHUR P. HURTER, JR., JOSEPH S. MARTINICH,
AND ENRIQUE R. VENTA*

The choice of a facility location and the design of the facility, including the choice of the quantity of each input to use, are intimately related. Consequently, one would expect facility location and facility design problems to be addressed as simultaneous problems in the theory of the firm. However, the traditional theory of the firm does not include spatial elements. Furthermore, the traditional theory of location, following Alfred Weber, does not include the choice of input mix. It is as if the developers of these traditional theories felt that a kind of separability was inherent in production-location problems. There are analytical advantages to the use of this separation, but it carries with it the possibility of suboptimization.

A number of authors have taken the point of view that production and location problems should be treated simultaneously. These efforts include the classic paper by Leon Moses and more recent papers by Hurter and Richard Wendell, David Emerson, A. J. Goldman, M. Bradfield, Noboru Sakashita, Robert Woodward, and Amir Khalili, Vijay Mathur, and Diran Bodenhorn. In one way or another, each of these authors contributes something to the question of the conditions under which the location decision and input-mix decisions can be treated separately. Our purpose here is to address this question directly, to draw together the contribution of these various authors, and to extend them.

I. Assumptions and Problem Formulation

A. *Assumptions*

1: *The facility produces a single product at rate* z_0.

2: *The facility uses* n *inputs at rates* $(z_1, \ldots, z_n)$.

3: *Input prices, at their sources, are known and fixed at* $p_j, j=1, \ldots, n$.

4: *The location of the input sources are known and fixed at* $(a_{1j}, a_{2j}) = a_j$.

5: *One facility is to be located at* $(x_1, x_2) = x$.

6: *Transportation costs for each input are proportional to distance and quantity at* r_j *(for example, \$/ton-mile).*

7: *Any* l_p *distance measure can be used. The distance from the facility to the source of input j is* $d(x, a_j)$.

8: *Customers pay for the cost of transporting product.* (Relaxed for the third and fourth problems.)

9: *The objective is to "design" a facility to produce output at a given rate* $\bar{z}_0$.

10: *Inputs and outputs are related by a production function* $z_0 = f(z_1, \ldots, z_n)$.

11: *Customer demand for the product is totally price inelastic.*

B. *First Problem*

The problem formulation suggested by this set of assumptions is that of choosing the facility location (x) and the input mix (z) so as to minimize production and transport costs while producing output at rate $\bar{z}_0$. This takes the form:

$$(1) \qquad \min_{x,z} \sum_{j=1}^{n} \left[r_j d(x, a_j) + p_j \right] z_j$$

subject to $\quad \bar{z}_0 = f(z_1, \ldots, z_n)$

---

*Professor of industrial engineering and management sciences and of finance, Northwestern University; assistant professor, School of Business Administration, University of Missouri–St. Louis; and instructor in management science, Loyola University, respectively. Work was partially supported through NSF grant ENG 77-07463.

This is a version of the problem addressed by Moses and later by Hurter and Wendell. Hurter and Wendell rewrote equation (1) introducing an "artificial" separation of the problem into two subproblems: one a minimization over z; the other a minimization over x. They recognized that, for an arbitrary location $\bar{x}$, the first-order condition for the minimization over z is

$$(2) \quad r_j d(\bar{x}, a_j) + p_j - \mu(\bar{x}) f_j(z^*) = 0 \qquad j = 1, \ldots, n$$

The asterisk denotes the optimal values and $\mu(\bar{x})$ is the Lagrangian multiplier interpreted as the marginal cost of production at location $\bar{x}$. Hurter and Wendell noticed that, when each first-order condition was multiplied by z_j^* and the result summed over j, a new form of the objective function was obtained:

$$(3) \quad \sum_{j=1}^{n} \left[r_j d(\bar{x}, a_j) + p_j \right] z_j^* = \mu(\bar{x}) \left[\sum_{j=1}^{n} z_j^* f_j(z^*) \right]$$

The left-hand side of (3) has the same form as the objective in (1). This suggests substituting the right-hand side of (2) as an equivalent objective in the production-location problem. If the right-hand side of (3) were independent of the choice of $z_1, \ldots, z_n$, then separation of the choice of x from that of z would be possible without danger of suboptimization. When $[\sum_{j=1}^{n} z_j f_j(z)]$ is a constant K for all z such that $\bar{z}_0 = f(z_1, \ldots, z_n)$, the right-hand side of (3) is independent of z. Hurter and Wendell identified a class of production functions exhibiting this property, and Goldman has shown that this class is identical to the class of homothetic production functions developed by Ronald Shephard. Thus:

THEOREM 1: *If the production function is homothetic, then equation* (1) *can be solved by first solving for an optimal location* x^* *and then for the optimal input mix* z^*. *In other words, the traditional separation is valid for the first problem* (1) *if the production function is homothetic.*

C. *Second Problem*

A related question to that of separability is whether or not the optimal location varies with the level of output. Homothetic production functions have the property that the expansion paths they generate are straight lines, meaning that optimal input *ratios* $z_i^*/z_j^*, i, j = 1, \ldots, n$ are not altered as the rate of output changes as long as Assumption 3 holds. Thus:

THEOREM 2: *If the production function is homothetic, the optimal location,* x^*, *and the optimal input ratios,* z_i^*/z_j^* *related to equation* (1), *are independent of the level of output,* z_0, *and a single location is optimal for all levels of output.*

PROOF:

Let $\theta(z_0) \triangleq \{(z, x) | f(z) = z_0$, and (z, x) is optimal in equation (1) for $z_0\}$. Let z_0^1 and z_0^2 be arbitrary rates of output (>0). Shephard has shown that there exists some scalar $\lambda > 0$ (for f homothetic) and that for all z^1 with $z_0^1 = f(z^1)$, $f(\lambda z^1) = z_0^2$. If (z^1, x^1) is optimal for z_0^1 and (z^2, x^2) is optimal for z_0^2,

$$\sum_{j=1}^{n} \left[r_j d(x^1, a_j) + p_j \right] z_j^1 \leqslant \sum_{j=1}^{n} \left[r_j d(x^2, a_j) + p_j \right] (z_j^2/\lambda)$$

$$\text{or} \quad \sum_{j=1}^{n} \left[r_j d(x^1, a_j) + p_j \right] \lambda z_j^1 \leqslant \sum_{j=1}^{n} \left[r_j d(x^2, a_j) + p_j \right] z_j^2$$

Then, $(\lambda z^1, x^1) \in \theta(z_0^2)$. Similarly, it can be shown that $(1/\lambda z^2, x^2) \in \theta(z_0^1)$ or $(z^2, x^2) \in \lambda\theta(z_0^1)$. Finally,

$$\theta(z_0^2) = \lambda\theta(z_0^1) \triangleq \{(\lambda z^1, x^1) | (z_1^1, x^1) \in \theta(z_0^1)\}$$

An alternative form for the proof of Theorem 2 emphasizes the separability through the formulation of the cost function. Based on Shephard's work, Hurter and Venta have shown that the *cost function* $C(x, z_0)$, which includes the delivered costs of the optimal (for z_0) mix of inputs, can be written

$$C(x, z_0) = h(x)g(z_0) \tag{4}$$

providing the production function is homothetic. Thus, the location that minimizes cost for some arbitrary $\bar{z}_0$, minimizes cost for all values of z_0 as required by Theorem 2.

Shephard has shown that the production function is homothetic if and only if the cost function is separable in input prices and output level. Then, (4) is a proper representation if and only if the production function is homothetic. Consequently, Theorems 1 and 2 could be written to show that homotheticity of the production function in (1) is necessary and sufficient for separability and for independence of location and level of output.

D. *Third Problem*

Suppose Assumption 8 is relaxed and the firm is required to pay the transport costs for its product. Let $b_k = (b_{1k}, b_{2k})$ be the location of the kth customer, let q_k be the quantity required by the kth customer and r be the unit transport cost of the firm's product. Then the design problem can be written

$$\min_{x,z} \sum_{k=1}^{n} rd(x, b_k)q_k + \sum_{j=1}^{n} \left[r_j d(x, a_j) + p_j \right] z_j \tag{5}$$

$$\text{subject to} \quad \sum_{k=1}^{m} q_k = z_0 = f(z_1, \ldots, z_n)$$

This is equivalent to

$$\min_x \left[\sum_{k=1}^{m} rd(x, b_k)q_k + \left\{ \min_{z \ni z^0 = f(z)} \sum_{j=1}^{J} \left[r_j d(x, a_j) + p_j \right] z_j \right\} \right]$$

The subproblem in { } is similar to (1) and, using (3), it can be written as $\min \mu(x)$. Consequently, (5) can be written

$$\min_x \mu(x) + \sum_{k=1}^{m} rd(x, b_k)q_k \tag{5'}$$

which is a "pure" location problem. Once again, it is possible to separate the choice of optimal location from the choice of the optimal input mix when f is homothetic.

E. *Fourth Problem*

An equivalent form of (5'), which is convenient for studying the effect of changes in rate of output on location and on the optimal input mix may be written in terms of the cost function:

$$\min_x \sum_{k=1}^{m} rd(x, b_k)q_k + h(x)g(z_0) \tag{6}$$

This function is not independent of the level of output z_0, nor of the quantities demanded by individual customers, q_k. Thus, when the costs of transporting products to market are borne by the producer and are of such a magnitude that they cannot be ignored, the optimal location x^* is *not*, in general, independent of the rate of output z_0, even when the production function is homothetic.

However, if the production function exhibits constant returns to scale, then $h(x)$ in (6) is the average cost of production and transportation of all inputs, which is a constant with respect to the level of production z_0 at location x. Then, (6) becomes

$$\min_x \sum_{k=1}^{m} rd(x_k, b_k)q_k + h(x)z_0 \tag{6'}$$

When z_0 is changed to $\beta z_0 (\beta > 0)$, and each q_k is changed to βq_k, the location that was optimal when output was z_0 will be optimal for output βz_0 for all $\beta > 0$.

II. Discussion

Traditional location analysis, which abstracted from the determination of the optimal input mix for the production of a given rate of output, was on firm ground. Whenever the production function is homothetic, the choice of an optimal location can be treated independently of the choice of an optimal input mix. Similarly, under these conditions, the theory of the firm was justified in abstracting from spatial considerations. Further, when the cost of transporting the product can be ignored, the same location is optimal for all levels of output.

Almost all the production functions commonly used in economic analysis, including many based directly on engineering design relationships, are homothetic or can be accurately approximated by homothetic functions. Consequently, the restriction of these results to homothetic production processes is only a modest limitation on their general applicability.

REFERENCES

M. Bradfield, "A Note on Location and the Theory of Production," *J. Reg. Sci.*, Aug. 1979, *19*, 263–66.

D. L. Emerson, "Optimum Firm Location and the Theory of Production," *J. Reg. Sci.*, Dec. 1973, *13*, 335–45.

A. J. Goldman, "Fixed Point Solution of Plant Input/Location Problems," *NBS J. Res.–B. Math. Sci.*, Apr./June 1974, *78B*, 79–94.

A. P. Hurter, Jr. and R. E. Wendell, "Location and Production—A Special Case," *J. Reg. Sci.*, Aug. 1972, *12*, 243–47.

_____ and E. R. Venta, "Production Function, Cost Function and the Spatial Theory of the Firm," mimeo., 1980.

A. Khalili, V. K. Mathur, and D. Bodenhorn, "Location and the Theory of Production: A Generalization," *J. Econ. Theory*, Dec. 1974, *9*, 467–75.

L. N. Moses, "Location and the Theory of Production," *Quart. J. Econ.*, May 1958, *72*, 259–72.

N. Sakashita, "Production Function, Demand Function and Location Theory of the Firm," *1967 Papers Reg. Sci. Assn.*, Vol. 20, 109–22.

Ronald W. Shephard, *Theory of Cost and Production Function*, Princeton 1970.

R. S. Woodward, "The Iso-Outlay Function and Variable Transport Costs," *J. Reg. Sci.*, Dec. 1973, *13*, 349–55.

Unique Equilibria, Pure Profits, and Efficiency in Location Models

By DENNIS R. CAPOZZA AND ROBERT VAN ORDER*

Recent papers in spatial microeconomics have analyzed the distortions that arise when space influences a firm's decision (for example, see our 1977a,b, 1978 papers; B. Curtis Eaton and Richard G. Lipsey, 1976b, 1978; H. S. Grace; M. L. Greenhut; Harold Hotelling). As this research has evolved, it has become clear that what might appear to be somewhat narrow issues in spatial economics are having rather general implications for economic theory as a whole. Excellent reviews have recently been provided by Eaton-Lipsey (1976b) and Greenhut. For example a number of undesirable features of the perfectly competitive model disappear when space is introduced. Demand curves faced by firms need not be horizontal, the size of the firm is no longer indeterminate, and increasing returns to scale need not preclude an equilibrium. The link between the perfectly competitive model and spatial models has been outlined in our 1978 paper.

As pointed out in a series of articles by Eaton and Lipsey (1976a,b, 1978), most of these models determine long-run equilibrium by assuming entry until profits are zero. For reasons to be discussed shortly, this may not be the right assumption. The long-run level of profits should be *derived*. It may be that long-run equilibrium is not unique and profits are nonzero.

This paper has two purposes. The first is to provide an exposition of the dynamics of entry in location models in some simple but interesting cases. The analysis is intended to complement and clarify work by Eaton and Lipsey. Important cases are presented in which profits are not zero, equilibrium is not unique, and history matters forever. The second purpose is to analyze welfare implications of the model. It is suggested that pure profit equilibrium may, in fact, be preferable to zero profit equilibrium.

In the next section we discuss the assumptions used here as well as in the Eaton-Lipsey papers. Section II develops a simple model with portable firms. Equilibrium is unique and economic profits are zero. Section III extends the model to the case of immobile firms. We find both pure profits and multiple equilibria. Section IV considers the consequences of multiple equilibria and pure profits.

I. Assumptions

Here we discuss assumptions common to location models.

ASSUMPTION 1: *Market.* The topology of the market is not crucial to the issue at hand; however, there are two obfuscations that we wish to avoid, both of which arise in a bounded market. First, in a bounded market the outermost firms behave differently from interior firms. The outer firm has an incentive to locate closer to an interior firm than to the market boundary, while interior firms generally prefer to be equally spaced. This is the Hotelling problem, which has been analyzed extensively. We avoid the problem by assuming the market to be unbounded, so that there are only interior firms.

The second problem arises when the market is finite in length (i.e., closed). Pure profits can arise for reasons unrelated to the central issue. Consider Paul Samuelson's volcanic island where firms are spaced around the perimeter of a circle. Let the market be uniform in density, firms have only fixed costs, and price be fixed and unitary. Assume that each firm requires a market of length Δ to break even. If the perimeter is L miles long and $L=N\Delta$, then N firms will exist and earn zero profits. But if the market is slightly larger so that $L=(N+\delta)\Delta$, where $0<\delta<1$, only N firms can survive profitably. The extra market length $\delta\Delta$ will be divided among the existing firms and each will earn pure profits equal to $\delta\Delta/N$.

---

*University of British Columbia and U.S. Department of Housing and Urban Development, respectively. An earlier version of this paper was presented at the Pacific Regional Science meetings in August 1977. We have benefited from the helpful comments of B. Curtis Eaton. Remaining errors are our responsibility, of course.

Notice, however, that this source of pure profits disappears in a large market. As market length L approaches infinity, the number of firms N also approaches infinity. Profits per firm fall to zero even if δ is close to its upper bound of one.

The above is clearly a result of indivisibilities and is a possible source of pure profits, but one which is less important in the typical large market. Since we wish to concentrate on other aspects of the issue, we assume the market to be unbounded and infinitely extensible. In addition, we assume a one-dimensional market for simplicity. This suppresses the market-shape issue of Eaton-Lipsey (1976).

ASSUMPTION 2: *Demand.* The pure profits issue does not appear to be sensitive to the assumption about individual demand curves. We make the simplest assumption of an all or nothing demand, where

$$q=\begin{array}{r|l} 1 & \text{if } p^d\leqslant p_0 \\ 0 & \text{if } p^d>p_0 \end{array}$$

where q is quantity demanded, p^d is delivered price, and p_0 is maximum price. Consumers buy one unit from the producer with the lowest delivered price. To prevent any producer from becoming a local monopolist, we also must have for $i=1,2$

$$p+tR_i\leqslant p_0$$

where p is mill price, R_i is distance to the ith competitor, and t is transport cost per unit distance.

ASSUMPTION 3: *Pricing Policy.* Firms are assumed to charge that price which maximizes profits.

ASSUMPTION 4: *Conjectural Variation.* The assumption made with respect to competitors' reactions or "conjectural variation" is critical and requires some elaboration. In the typical location model, equilibrium price and market radius are determined from two relationships. The first relationship between market radius and optimal price is obtained by maximizing profits with respect to price. The second is obtained by assuming that equilibrium occurs when profits are zero; that is, firms enter as long as profits of existing firms are positive (see our 1978 paper).

Presumably new firms enter because they expect *their* profits to be positive, not because existing firms have positive profits. In most models, expected profits of new firms and observed profits of existing firms are identical. If entering firms are "small" relative to the market, and if entering firms face the same cost and demand curves as existing firms (which is the usual assumption), then entering firms should believe that they will earn as much as existing firms. Hence, normally, *observed* profits or losses can be taken as the stimulus to entry or exit.[1]

However, in locational models, potential entrants are not necessarily the same as existing firms. Since they must locate between two old firms, their market may only be about half the size of existing firms' markets. Hence, their *expected* profits could be negative while existing firms' current profits are positive. If existing firms hold their positions, entering firms cannot expect to garner as large a market. Eaton-Lipsey have labelled the assumption that firms expect other firms to stay put after entry "zero locational conjectural variation" (*ZLCV*). It is this assumption and its alternatives on which we focus our analysis.

Zero conjectural variation (*ZCV*) is an extension of the usual notion in monopolistic competition that individual firms face a large number of competitors. We believe that *ZCV* is plausible though controversial as far as price is concerned (*ZPCV*), (see our 1978 paper). Fortunately, it is not essential, so we need not take issue with it except to note that the results are consistent with other price conjectural variations (see fn. 3 below).

---

[1] See Eaton and Lipsey (1978) for a similar argument.

Following the Eaton-Lipsey (1978) model, we distinguish two types of equilibrium:

1) Fixed-numbers equilibrium: when no change in price or location will increase profits.

2) Free entry equilibrium: when expected profits of new entrants are nonpositive.

For the purposes of dynamics, the central issue concerns fixed-numbers equilibrium. In Eaton-Lipsey (1976a; 1978) it is used mainly as a condition of long-run equilibrium, but in the simulation routine in Eaton it is assumed to occur in the short run; that is, when a new firm enters existing firms move to new locations to establish a limited equilibrium. For this to occur firms must be portable.

In building a model of long-run equilibrium, it is desirable to have firms believe only things that are true. If we assume fixed-number equilibrium to hold, then if a new firm enters, neighboring firms do *not* stay put. They move away, and force their neighbors to move away, and so on. Indeed, one firm entering on an infinite line will affect neither the neighboring firms' prices nor the distance between firms since firms will spread out evenly in equilibrium. In the long run, entering firms should be aware of this. Rather than expecting a zero variation, they should expect a positive locational response; that is, they should expect that firms will move when they enter and that distance between firms will remain constant *as long as firms are portable*. If firms are not portable, we cannot have fixed-numbers equilibrium in the short run. Thus, there may be a conflict in assuming both fixed-numbers equilibrium and *ZLCV*.

The way out, of course, is to assume that there are costs of moving firms. This tack is followed by Eaton-Lipsey (1978). It is certainly a reasonable approach, but again it implies that we cannot have fixed-numbers equilibrium with respect to location; that is, the costs of relocating will hinder movement to equilibrium in the short run. We consider two polar cases. In the first, firms are portable and believe (correctly) that their entry will cause other firms to move. On the infinitely long line there is no post-entry change in market area. In the second case, firms are "nailed down" and recognize their own and other firms' immobility. In both cases firms assume no price reactions by rivals.

ASSUMPTION 5: *Linearity*. To simplify matters we assume that total costs are linear in output and that transport costs per good-mile are a constant, t.

II. Equilibrium with Portable Firms

Assume that population density is constant and equal to one person per unit area, and that each person buys one unit of the good in question. With demand perfectly inelastic, the only decision for consumers to make concerns the store from which they buy. It will be the store with the lowest delivered price. The boundary between two firms, given their mill price, will be that point where delivered prices of each firm are equal. If a firm raises its price, this boundary contracts and the firm loses customers to competitors. Hence, the slope of the firm's demand curve comes from changes in market area rather than changes in the quantity purchased by each individual.

Instead of assuming that firms are completely portable, a rather unrealistic assumption, we assume that firms are nailed down for one period; but at the end of the period they can relocate as freely as new entrants. Hence, during the period we have both *ZLCV* and *ZPCV*, but between periods new firms enter if profits of existing firms are positive. It is then straightforward to show that[2]

$$Q = 2R + \frac{1}{t}(\bar{p} - p) \qquad (1)$$

---

[2] This is a variant of the demand curve in our 1977a paper. It is derived by equating delivered price lines to

where $2R$ is the distance between the firm's competitors, t is unit transport costs, p is the firm's price, $\bar{p}$ is the average price of its two nearest competitors, and Q is the demand faced by the firm.

Let total costs be given by $k+cQ$, so that profits are

$$Y=(p-c)Q-k \tag{2}$$

Then profit maximization gives[3]

$$p=\frac{1}{2}(c+\bar{p}+2tR) \tag{3}$$

Since firms are identical, they should all charge the same price in equilibrium. Equating p and $\bar{p}$, we obtain "short-run" equilibrium price as a function of the predetermined density of firms; i.e.,

$$p=c+2tR \tag{4}$$

It should be noted here that the location of the individual firm between its rivals is indeterminate; in particular, it is just as profitable to be next to one of its rivals as it is to be located exactly between them.[4] Only market diameter D matters, the breakdown into distance to the left-hand and right-hand firms being of no consequence. This results from the inelastic demand curve. If demand had some price elasticity, firms would always choose to be in the center of the market (see Eaton-Lipsey, 1978).

For reasons discussed in the previous section, let entry with portable firms follow the usual mechanism in which entry depends on actual profits. New firms expect to earn exactly the same profits as existing firms. Entry ceases when profits of existing firms are zero. Hence, in the long run, the left-hand side of (2) is set equal to zero, and the zero profit price is found to be

$$p=c+k/2R \tag{5}$$

Equations (4) and (5) are two equations in p and R and yield the long-run levels of both. The resulting long-run levels of p and R in this mobile firm case are

$$p_M^*=c+(kt)^{1/2};\ R_M^*=\left(\frac{k}{4t}\right)^{1/2}$$

This equilibrium is unique, and profits are zero; that is, if there are no costs to changing locations, there are neither pure profits nor multiple equilibria. Firms base entry decisions on profits of existing firms and enter until profits are zero for all firms.

The model here can be interpreted either as one of portable firms (between periods) or as one in which firms last only one period. In the latter interpretation, all firms leave the industry, and a new set of firms takes over at the end of each period. This interpretation raises the question of gradual depreciation.

If capital depreciates gradually over several periods, an entering firm will realize that in the long run some rivals will leave. However, if there is a cost to waiting, entrants must expect positive long-run profits before they will enter and sustain temporary losses. We bypass this messy intermediate case and consider the other extreme where capital is both immobile and indestructible.

Note that it is not the case, as is sometimes argued (for example, see Eaton-Lipsey, 1978, and Nicholas Kaldor), that indivisibilities of capital or scale economics prevent zero profit equilibrium. The simple model here has both. Rather it is the immobility of capital that causes it.

III. Equilibrium with Immobile Firms

The dynamics behind the analysis of the preceding section is deliberately simple. Each time new firms enter, all firms relocate

---

find the boundary between firms and then noting that with unit demand per unit area, total demand is simply equal to total market area. The only difference here is the average price term $\bar{p}$, which arises because firms are not assumed to charge the same price.

[3] Maximizing (2) with respect to price gives

$$\partial Y/\partial P=0=(p-c)\,dQ/dp+Q$$

with $ZPCV$ $(d\bar{p}/dp=0)\,dQ/dp=1/t$ and (3) follows immediately. Other price conjectural variations can be treated similarly with no qualitative difference in the results that follow.

[4] We are indebted to Eaton for pointing this out to us.

to establish limited equilibrium. Because of the inelastic demand, firms are indifferent to where they locate; they can be either half-way between neighboring firms or asymmetrically spaced. They need not be equidistant. If we assume that firms are immobile, the actual spacing does matter since firms understand that there will not be continued relocation. They will consider existing locations given and will locate where the market is largest. The problem of firms entering when market possibilities are of uneven size is unnecessarily complicated. Since our immediate goal is to provide a plausible example of location dynamics and to establish that the solution is not unique, we simply assume that firms are evenly spaced and always enter at the center of the market on the grounds that the general character of our results will probably hold in more complicated cases.[5] We do assume fixed-numbers equilibrium for price; that is, once firms have entered, market price is given by (4).

If firms are evenly spaced, equation (4) can still be applied easily. The entering firm is now forced to locate between two immobile firms. It's radius will be half that of existing firms. Letting $\hat{R}$ be the expected radius of new firms, we have

$$\hat{R}=\frac{1}{2}R \tag{6}$$

where R is now redefined to be the actual market radius of existing firms before entry. For entry to stop, we must have zero profit for newcomers. By the *ZPCV* assumption, firms are expected to hold their prices constant. From (3) and (6), the profit-maximizing price for new entrants in terms of current distance between firms is

$$p=\frac{c+tR+\bar{p}}{2} \tag{7}$$

We know, from (4), that

$$\bar{p}=c+2tR \tag{8}$$

so that

$$p=\frac{c+tR+c+2tR}{2}=c+\frac{3}{2}tR \tag{9}$$

gives the price new firms intend to charge under *ZPCV* and *ZLCV* given the radius. Expected profits at this price are

$$\hat{Y}=2(p-c)\frac{1}{2}R-k$$

and using (9)

$$\hat{Y}=\frac{3}{2}tR^2-k \tag{10}$$

Then entry will stop when expected profits for new firms are zero and the actual distance between these immobile firms (denoted R_I^*) is

$$R_I^*=\left(\frac{2k}{3t}\right)^{1/2}>R_M^*=\left(\frac{k}{4t}\right)^{1/2}$$

The price associated with R_I^* is

$$\hat{p}_I=C+\left(\frac{3}{2}kt\right)^{1/2}>p_M^*=c+(kt)^{1/2}$$

The model is depicted in Figure 1. If the current radius is greater than R_I^*, say R_1, with existing firms charging P_1, new firms will enter charging P_2, forcing all firms to cut their price and decreasing market sizes by half. This process will continue until a radius R_I^* or less is reached. At this point additional new firms, who would charge $\hat{p}_I$ or less if they entered, do not enter, and the existing firms charge a market price less than or equal to

$$p_I^*=C+\left(\frac{8}{3}kt\right)^{1/2}>p_M^*$$

The actual market price depends on that radius between zero and R^{**} on which the system comes to rest. Any initial position to the left of R_I^* will be maintained. Long-run equilibrium can be anywhere along the line cB in Figure 1. Existing firms' profits will be positive for $R_I^*<R<R_I^*$, but negative for $R<R_M^*$.

---

[5] Again, if individual demand curves have the usual slope, location will always be at the center.

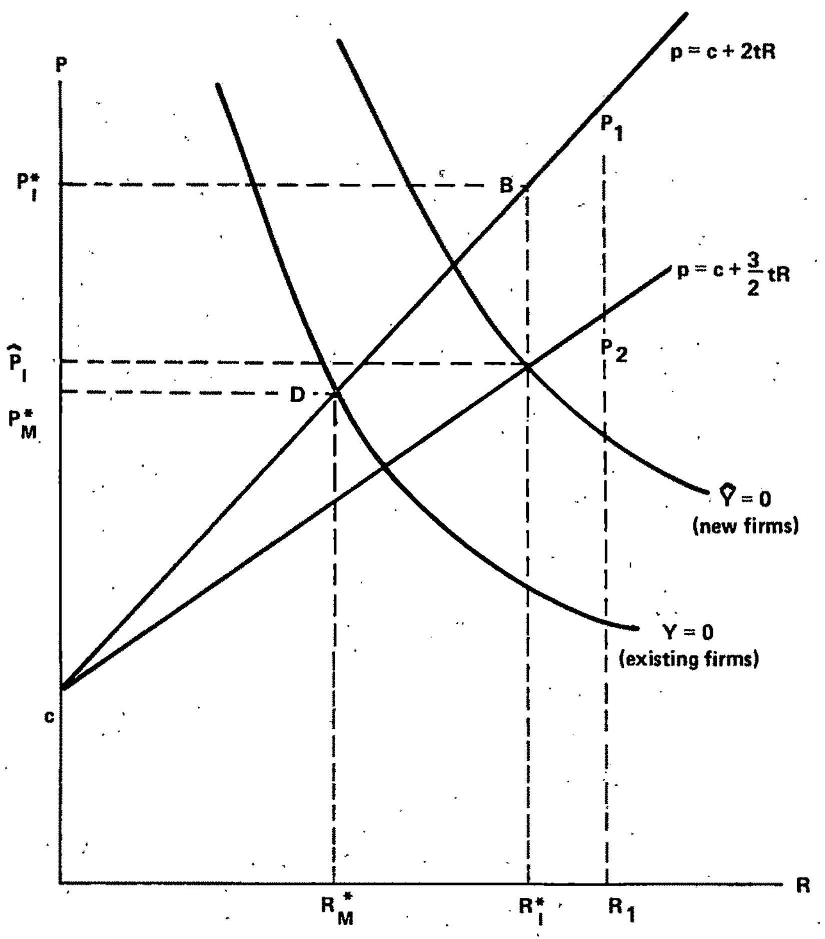


FIGURE 1. LONG-RUN EQUILIBRIUM WITH IMMOBILE FIRMS

This is essentially the Eaton-Lipsey result, depending on costs of shifting capital as well as indivisibilities; that is, the equilibrium price and radius are indeterminate unless the initial conditions are specified (multiple equilibria), and profits need not be zero.

An additional complication can arise here. In the portable model it made sense to argue that new entrants would not affect other firms' prices, since the firms moved easily. Here, when a firm enters between two immobile firms it affects their price, and it should expect to, undermining the *ZPCV* assumption. One could, however, simply assume that firms enter whenever the postentry market radius is large enough to yield profits at the current price; that is, when $p - c > k/2\hat{R} = k/R$. Roughly similar results would be obtained.

Alternately, letting $\pi = d\bar{p}/dp$, the conjectural variation, (4) above becomes $p = c + 2tR/(1-\pi)$ and similar results follow. An interesting special case is the "rational" one in which π is not just conjectured but "true." It can be shown that in that case $\pi = 1/3$ (see fn. 3).

IV. Further Observations and Welfare Implications

Comparative statics in the immobile case is complicated, depending on whether or not we start to the left of R_I^*, and on the direction of the change. To the left of R_I^*, changes in fixed costs k, or in population, will (unlike in the portable case) not affect price, but (as in the portable case) increases in marginal cost c or transport cost t will raise it. If we are at R_I^*, a rise in k will work qualitatively as it does in the portable case, but a decline will not. Changes in c and t will be similar to the portable case. Hence, comparative statics results of the two models are broadly similar.

Of some interest in models like these is the effect of the system on resource allocation. In our model, this problem has only one dimension. Do we minimize average costs?

Total transportation costs are given by

$$TTC = tR^2$$

and production costs are

$$TPC = k + 2Rc$$

Then average total costs are

$$ATC = \frac{k}{2R} + \frac{1}{2}tR + c$$

The optimal radius given by minimizing ATC is

$$R^o = \left(\frac{k}{t}\right)^{1/2} > R_I^* > R_M^*$$

In both models there are too many firms, leading to suboptimal resource allocation.

In the immobile case, the radius can be anywhere from zero to R_I^*. Suppose, however, that the basic parameters c, k, t, etc. have been constant. Let equilibrium be the result of a process which began with excess profit and a radius greater than R_I^*. Then the actual radius might be close to R_I^*, which is closer to R^o than is R_M^*. Thus, the seemingly inferior situation where entry is blocked because of immobile capital may be closer to optimal than the case where resources are free to move.

One interesting question to ask is whether the above market failure is a result of the incorrect expectations of new entrants; that is, would correct foresight with respect to postentry market radius and price result in an optimal allocation?

Suppose entrants believe that if they enter at a midpoint between firms, other entrants will fill in all other midpoints. Then every market will be cut in half. In this situation entry will only be profitable if the pre-entry radii are more than twice the break-even radius R_M^*. Equilibrium is sustainable for $0 \leqslant R \leqslant 2R_M^*$.

Since we have

$$R_M^* = \frac{1}{2}\left(\frac{k}{t}\right)^{1/2}$$

the optimal radius is

$$R^o = \left(\frac{k}{t}\right)^{1/2} = 2R_M^*$$

Therefore if the system starts from a point to the right of $2R_M^*$ at $4R_M^*$, the next round of entry will bring the system to $2R_M^* = R^o$, the optimal radius. Any other starting point will result in a suboptimal allocation. Thus with the above foresight there is a chance that the allocation will be optimal if initial conditions are fortuitous. In general there is still an efficiency loss. The interesting "second best" aspect of the model is retained: *Spatial competition is inefficient because there are too many firms; but having a barrier to entry in the form of large adjustment costs may improve efficiency.*

A further case of interest is that where fixed capital is portable but infinitely durable and not shiftable to other industries. In such a model we cannot have excess profit equilibrium, but we can have negative profit equilibrium. In Figure 1 equilibrium can be anywhere along the segment cD. Only if the capital depreciates rapidly, can we have the unique zero profit equilibrium.[6]

V. Conclusions

This paper analyzes several location models and tries to show that the positive profits/multiple equilibria result is plausible.

---

[6] This case was pointed out to us by Eaton.

Indivisibilities, capital mobility, and capital durability play a central role in the models. Indivisibilities alone will cause positive profits if the market is finite in size and small relative to the break-even size of the firm. In the more typical case where the total market is large relative to firm size, indivisibilities are necessary but not sufficient. Some degree of capital immobility and durability is required as well. Finally, while long-run equilibrium can be expected to be "inefficient," the pure profit equilibrium may entail higher welfare levels than zero profit equilibrium.

REFERENCES

M. J. Beckmann, "Spatial Cournot Oligopoly," *1972 Papers Reg. Sci. Assn.*, Vol. 28, 37–47.

D. R. Capozza and R. Van Order, (1977a) "A Simple Model of Spatial Pricing Under Free Entry," *Southern Econ. J.*, Oct. 1977, *44*, 361–67.

______ **and** ______, (1977b) "Pricing Under Spatial Competition and Spatial Monopoly," *Econometrica*, Sept. 1977, *45*, 1329–38.

______ **and** ______, "A Generalized Model of Spatial Competition," *Amer. Econ. Rev.*, Dec. 1978, *68*, 896–908.

B. C. Eaton, "Free Entry in One-Dimensional Models: Pure Profits and Multiple Equilibria," *J. Reg. Sci.*, Apr. 1976, *16*, 21–33.

______ **and R. G. Lipsey,** (1976a) "The Non-Uniqueness of Equilibrium in the Löschian Location Model," *Amer. Econ. Rev.*, Mar. 1976, *66*, 77–93.

______ **and** ______, (1976b) "The Introduction of Space into the Neo-Classical Model of Value Theory," in Michael J. Artis and A. R. Nobay, eds., *Studies in Modern Economic Analysis; the Proceedings of the Association of University Teachers of Economics*, Edinburgh; Oxford 1976, 59–96.

______ **and** ______, "Freedom of Entry and the Existence of Pure Profit," *Econ. J.*, Sept. 1978, *88*, 455–69.

H. S. Grace, Jr., "Professor Samuelson on Free Enterprise and Economic Efficiency: A Comment," *Quart. J. Econ.*, May 1970, *84*, 337–40.

M. L. Greenhut, "Impacts of Distance on Microeconomic Theory," *Manchester Sch. Econ. Soc. Stud.*, Mar. 1978, *46*, 17–40.

______, **M. Hwang, and H. Ohta,** "Observations on the Shape and Relevance of the Spatial Demand Function," *Econometrica*, July 1975, *43*, 669–82.

H. Hotelling, "Stability in Competition," *Econ. J.*, Mar. 1929, *39*, 41–57.

N. Kaldor, "Market Imperfection and Excess Capacity," *Economica*, Feb. 1935, *2*, 33–50.

P. A. Samuelson, "The Monopolistic Competition Revolution," in Robert E. Keunne, ed., *Monopolistic Competition Theory: Studies in Impact*, New York 1967, ch. 5.

______, "Professor Samuelson on Free Enterprise and Economic Efficiency: Reply," *Quart. J. Econ.*, May 1970, *84*, 341–45.

Harold Hotelling and Marginal Cost Pricing

By EUGENE SILBERBERG*

Perhaps the classic theorem of taxation in relation to welfare economics is the proposition that if all commodities are taxable, then an "income" (lump sum) tax will leave consumers on a higher indifference level than a system of excise taxes which yield prices that are not proportional to marginal costs. The proof of this proposition is attributed to Harold Hotelling's paper, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." Paul Samuelson cites Hotelling's paper as the *pièce de résistance* of Hotelling's work in economics.[1] Although the principle of marginal cost pricing appears in earlier works (dating back at least to Jules Dupuit), Hotelling, according to Samuelson, "was the first to generalize the argument from a Marshallian partial equilibrium analysis of a single market to the rigorous analysis of interrelated goods" (p. 24).

The theorem is, of course, subject to certain well-known qualifications, some of which were pointed out by Hotelling himself. It is difficult to imagine truly nondistorting, lump sum taxes. Taxes on income distort the labor leisure choice. Also, what are commonly called "income taxes" are in fact complicated rules taxing income from different sources differently, and in the case of taxing interest on savings, a tax on future consumption at higher rates than present consumption.

The purpose of this note is to demonstrate that, completely aside from the above qualifications and the other merits of his paper, Hotelling did *not* prove the above stated theorem. Hotelling's mathematics is in fact a simple argument about revealed preferences. It was an error of logic on Hotelling's part to apply the revealed preference analysis to the marginal cost pricing theorem.

The theorem was also discussed by Milton Friedman who correctly criticized the "proofs" offered by M. F. W. Joseph and others as having committed a fatal error in logic. Noting that Joseph's proof in no way involved having prices equal to marginal cost originally, the result had to be independent of whether or not prices before taxes were set at marginal cost. Friedman cautions that arriving at a correct conclusion does not render a proof logically valid. Hotelling's proof, however, was said to be valid, precisely because Hotelling "takes account of conditions of cost of production" (pp. 101–02).

Following Samuelson, the theorem follows simply from maximizing a utility function $U(q_1,\dots,q_n)$ of all commodities $q_1,\dots,q_n$, subject to a production transformation function $g(q_1,\dots,q_n)=0$. Taking the first partials of the Lagrangean $L=U+\lambda g$ yields the conditions,

$$U_i/U_j=g_i/g_j \qquad i,j=1,\dots,n$$

Thus, at the utility maximum, consumers' marginal evaluations of each good must equal the respective marginal cost of that good, expressed in terms of some numeraire commodity. Assuming that all commodities are taxable, any system of taxes which leaves $p_i/p_j \neq MC_i/MC_j$ where p_i=price (to consumers) of good i is necessarily not a Pareto optimum. If some item in the economy is fixed in supply, then taxing that commodity leaves the consumption bundle unchanged and the proceeds can be used to finance the distribution of some public good at $p_i=MC_i=0$. Thus, "lump sum" taxation is superior, in the Pareto sense, to nonproportional excise taxation.[2]

---

*Professor, University of Washington. I have benefited from discussions with Yoram Barzel and Robert Halvorsen, and remarks by an anonymous referee.

[1] Interestingly enough, Samuelson chose to substitute his own proof of the theorem, rather than reproduce Hotelling's argument.

[2] With no savings, a tax structure which raises all prices by the same proportion (not necessarily the same

Hotelling, however, did not present this line of argument. His argument for $p_i = MC_i$ is based on verbal discussion and diagrams in a partial equilibrium context. He then proceeds to a mathematical discussion that, as is shown below, is quite irrelevant to the issue under discussion. For ease of comparison, I will number the equations as they are numbered in Hotelling's paper.

I. Hotelling's Proof

Hotelling begins his argument by assuming initially, $p_i = MC_i$, $i = 1, \ldots, n$; i.e., that all prices are set at the marginal costs of production. The symbol m represents income after payment of some lump sum income tax, resulting in the budget condition

$$\sum p_i q_i = m \tag{4}$$

The q_i chosen are assumed to be those that maximize some utility function $\Phi(q_1, \ldots, q_n)$ subject to (4). Therefore, if any *other* q_i, say $q_i' = q_i + \delta q_i$, are chosen, then necessarily the consumer is at a less preferred position. That is,

$$\sum p_i q_i' = m \tag{5}$$

implies

$$\Phi(q_1, \ldots, q_n) > \Phi(q_1', \ldots, q_n') \tag{7}$$

However, (4) and (5) together imply

$$\sum p_i \delta q_i = 0 \tag{6}$$

That (6) implies (7) is nothing other than the definition of revealed preference, as was being developed by Samuelson concurrently with Hotelling's article. It is the revealed *preference* idea which leads Hotelling, incorrectly, to his conclusion: Let $\delta p_i = p_i' - p_i$ represent the excise tax (or subsidy) on commodity i resulting in the purchases $q_i' = q_i + \delta q_i$. Let $\delta m = m' - m$ represent the reduction in income taxes available due to the imposition of the excise taxes.

The individual's budget constraint is now $\sum p_i' q_i' = m'$, or

$$\sum (p_i + \delta p_i)(q_i + \delta q_i) = m + \delta m \tag{10}$$

Subtracting equation (4) yields

$$\sum p_i \delta q_i + \sum \delta p_i (q_i + \delta q_i) = \delta m$$

or

$$\sum p_i \delta q_i = \delta m - \sum \delta p_i (q_i + \delta q_i) \tag{11}$$

It follows from equation (11), and the revealed preference argument above ((6) implies (7)), that if the income tax revenue δm equals the excise tax revenue $\sum \delta p_i (q_i + \delta q_i)$ that the consumer is necessarily worse off by the replacement of the income tax by excise taxes which change relative prices.[3] Hotelling concluded from this algebra that

> ...if we start from a system of excise taxes, *or any system in which sales are not at marginal costs,* this argument shows that there is a possible distribution of personal income taxes such that everyone will be better satisfied to change to the system of income taxes with sales at marginal cost.
> [p. 292, emphasis added]

But how can this be a valid conclusion from the above analysis? Equations (4)–(11) are an argument about revealed *preferences*. The condition $p_i = MC_i$ is nowhere used. Indeed, if one assumed that initially $p_i \neq MC_i$ and that the δp_i were chosen so as to make $p_i' = p_i + \delta p_i = MC_i$, the same logic would hold, "proving" that the consumer is worse off at q_i', where prices now equal MC!

In words, what Hotelling did, starting with his equations (4) and (5), was to begin with

---

tax rate on all commodities) is equivalent to an income tax. If the supply of income, for example, labor, is in perfectly inelastic supply, then this proportional taxation on the remaining commodities is in effect a lump sum tax.

[3] From revealed preference considerations, the result follows if the income tax revenue is less than or equal to the excise tax revenue.

a consumer who has achieved a point of maximum utility subject to a budget constraint. He then moves the consumer to some other point on the budget hyperplane, through changes in excise and income taxes. This point is necessarily less preferred than the original choice: If a consumer chooses one point on a budget constraint as his utility maximum, then any other point (ignoring multiple equilibria) on that constraint must be less preferred. The condition $p_i = MC_i$, $i = 1,\ldots,n$ is irrelevant to this argument.

Hotelling arrived at his conclusion on taxation by *appending* to the above argument, that, *if* initially $p_i = MC_i$ for every commodity, *then* a system of distortions will lead to a less preferred position. But this surely does not prove that maximum utility subject to a resource constraint *must* occur at $p_i = MC_i$, $i = 1,\ldots,n$. The resource constraint, the production transformation surface, is nowhere used in Hotelling's equations. Hotelling just showed that *if* the utility maximum occurred at $p_i = MC_i$, $i = 1,\ldots,n$, *then* since a system of compensated distortions would lead the consumer away from that point, *those* distortions would lead to a less preferred point. But this proposition is very far indeed from the important theorem Hotelling correctly conjectured.

Hotelling restates his "fundamental theorem" as:

> *if a person must pay a certain sum of money in taxes, his satisfaction will be greater if the levy is made directly on him as a fixed amount than if it is made through a system of excise taxes which he can to some extent avoid by rearranging his production and consumption.*
>
> [p. 293]

This statement of the "theorem" would be correct if the term "excise taxes" were to mean *any* change in the original (no tax) relative prices faced by the consumer. However, if the original prices did not equal marginal cost for reasons of, say, monopoly, the same conclusion would hold, that is, consumers would prefer to have a given tax revenue extracted from them via an income tax rather than a system of nonproportional taxes on the commodities themselves. But this would not in general be a Pareto-superior move for the economy; excise taxes on the nonmonopolistically produced commodities would be required, as was pointed out by Friedman.[4]

II. Concluding Remarks

Hotelling in fact made the same error in logic as did the other writers cited by Friedman. The fact that if at the first bundle (the utility maximum) prices equal marginal cost, then a correct conclusion emerges about excise taxation which does not render the implied syllogism valid. It is *not*, in fact, a valid syllogism. It is impossible to derive from a revealed preference argument a conclusion about the efficiency of marginal cost pricing. If, initially, $p_i \neq MC_i$, it would still follow via Hotelling's reasoning that the second bundle was less preferred than the original, since the original bundle was the utility maximum and both points lie on the same budget constraint. But the second bundle could be any bundle at all (on the same budget constraint), including the point where $p_i = MC_i$ for all commodities. Hotelling's syllogism would then lead to a false conclusion; therefore, *it is an invalid syllogism*. Hotelling's paper of course remains a landmark, a seminal work in the theory of public goods and taxation. The issues raised in that paper are still under discussion, and the general problem of "optimal" taxation is as yet incompletely resolved. Perhaps it was the overall impact of the paper which obscured an avoidable error in logic over the past forty years.

---

[4] It is *not* possible to infer from any of the above that if "excise" taxes are already being levied, then consumers would prefer that any additional tax revenue be collected by income rather than excise taxes. Excise taxes that restored proportionality to the original relative prices would be preferred.

REFERENCES

J. Dupuit, "On the Measurement of the Utility of Public Works," in Kenneth J. Arrow and Tibor Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969.

M. Friedman, "The Welfare Effects of an Income Tax and an Excise Tax," in his *Essays in Positive Economics*, Chicago 1953.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," in Kenneth J. Arrow and Tibor Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969.

M. F. W. Joseph, "The Excess Burden of Indirect Taxation," *Rev. Econ. Stud.*, June 1939, *6*, 226–31.

P. A. Samuelson, "Harold Hotelling as Mathematical Economist," *Amer. Statistician*, June 1960, *14*, 21–25.

Uncertain Externalities, Liability Rules, and Resource Allocation: Comment

By VAROUJ A. AIVAZIAN AND JEFFREY L. CALLEN*

In a recent article in this *Review*, Peter Greenwood and Charles Ingene (hereafter G-I) deny the validity of the Coase Theorem when applied to an uncertain world that contains no stock market. Recall that the theorem maintains that resource allocation is neutral with respect to liability rules. The gist of their argument follows. Suppose firm *A*, whose profits are deterministic,[1] is legally permitted to pollute firm *B*, whose profits are stochastic. Since there are no transaction costs by assumption, firm *B* has the incentive to bribe firm *A* to reduce its polluting activities. Because firm *B* is liable for the damages, which are stochastic *ex ante*, the bribe will reflect firm *B*'s degree of risk aversion. If, on the other hand, firm *A* must legally recompense firm *B* for the damages, firm *A* will bribe firm *B* to reduce its operations and, consequently, lessen the damages. Since firm *A* is legally bound to insure the damages, in this case the bribe will reflect firm *A*'s degree of risk aversion. In the absence of a stock market which reconciles differential risk preferences, liability rules will have an asymmetrical impact on resource allocation. Therefore, in the presence of risk and the absence of transactions costs and a stock market for risk sharing, resource allocation is dependent on liability rules.

We will show that the G-I argument has little merit in a world of zero transaction costs. Just as the lack of a market for the externality does not preclude exchange (or bribery) in the standard Coase scenario, the absence of a stock market does not preclude risk sharing in an uncertain world. We will illustrate the G-I argument and our counterargument by an example.

Suppose firm *A*, the polluter, can produce one more unit of output at a marginal profit of \$50. The additional output has a negative impact on firm *B*'s profits. Let us first suppose that the damages are \$60 certain. Thus, if firm *A* is liable for the damages, the marginal unit is not produced. If firm *B* is liable, it bribes firm *A* and again the output is not produced. If the damages are only \$40 certain, a similar argument would show that, independent of the liability rule, the marginal unit is produced. Both cases illustrate the Coase Theorem at work.

Now, suppose the damages are uncertain, that is, distributed according to some known probability distribution. Firm *B* evaluates the potential damages at \$75 in certainty-equivalent terms. Firm *A* is less risk averse and evaluates the damages at \$40 in certainty-equivalent terms. Greenwood and Ingene would argue as follows. If firm *A* is not liable, firm *B* will try to bribe firm *A*. The minimum bribe firm *A* will accept is \$50 and the maximum bribe firm *B* will offer is \$75. Clearly, there are gains from trade and the marginal unit is not produced. If firm *A* is liable, the damages are valued at \$40, marginal profit is \$50, and the unit is produced. It would appear that the differential risk preferences vitiates the Coase Theorem.

We are able to counter the G-I argument by noting that firm *A* has the additional (and profitable) option to assume the liability when firm *B* is liable. In other words, firm *A* will write a (costless) contract to insure firm *B* against the damages for a mutually satisfactory fee of between \$40 and \$75. In fact, a bargaining solution which does not lead to an insurance contract is generally (Pareto) suboptimal. To see this, suppose that, in the absence of such a contract, firm *B* would have bribed firm *A* by the maximum \$75 so that the marginal unit would not be produced. But if firm *A* insures firm *B* for \$70 and produces the marginal unit, then firm *A* would gain \$80: \$50 in marginal profits plus the \$70 insurance fee minus \$40 for damages. Firm *B* also benefits by losing only \$70 instead of \$75. If, on the other hand, firm *A* would have accepted the minimum bribe of \$50, both firms could be made better off if the insurance fee is \$45. Firm *A* would gain \$55 and firm *B* would forfeit \$45 instead of \$50. Therefore, independent of the liability rule, the marginal unit is produced and the Coase Theorem holds.

---

*Faculty of Business, McMaster University. We wish to thank an anonymous referee for correcting an error in our numerical example.

[1] This assumption is made for simplicity of exposition. Clearly, both firms' profits could be stochastic.

Although we assumed above that firm *B* is more risk averse than firm *A*, a similar argument would prevail if firm *A* is more risk averse than firm *B*. Suppose that firm *A* evaluates the damages at \$75 and firm *B* at \$40 in certainty-equivalent terms. Greenwood and Ingene would argue that, if firm *B* is liable, the marginal unit is produced since firm *B* is unwilling to bribe more than \$40. If firm *B* is not liable, firm *A* will not produce the unit since the damages are evaluated at \$75 while marginal profits are only \$50. The Coase Theorem does not appear to hold. However, if firm *B* insures firm *A* for a fee of between \$40 and \$50, both firms can be made better off. Again, independent of the liability rule, the unit is produced.

Greenwood and Ingene also maintain that, even if there is a stock market, the Coase Theorem is not valid in the presence of indivisibilities: "An autonomous manager who receives a significant segment of his income from the firm will obey his own, not the market's, risk measure. Similarly, if '...some shareholder [holds] a large block of stock...' the firm should use his, not the market's, risk measure" (p. 303). Although the G-I arguments are compelling, the logic is faulty. If, in fact, all transactions are costless, so is monitoring the activities of the so-called autonomous manager. Therefore, even if his risk preferences (or preferences for perquisites) differ from that of his shareholders, the manager would be unable to effect his ambitions. In a transactionless world, the autonomous manager is a contradiction in terms.

A similar criticism applies to the dominant shareholder notion. In a world of zero transactions costs, contracts will be written for every contingency and state of nature so that markets will be complete. In such a world, a well-known unanimity theorem obtains, namely, that all shareholders will be unanimous about the firm's investment or production decisions.[2] Therefore, even if there is a dominant shareholder, his preferences and those of the other shareholders will coincide. Similarly, if two firms merge, the preferences of both sets of shareholders will coincide so that resource allocation is independent of who controls the corporation.

The upshot of our criticism of the G-I article is that the Coase Theorem is far more robust than they think, primarily because of the powerful assumption of zero transactions costs.

---

[2] See, for example, Steinar Ekern and Robert Wilson.

REFERENCES

S. Ekern and R. Wilson, "On the Theory of the Firm in an Economy with Incomplete Markets," *Bell J. Econ.*, Spring 1974, *5*, 171–80.

P. H. Greenwood and C. A. Ingene, "Uncertain Externalities, Liability Rules, and Resource Allocation," *Amer. Econ. Rev.*, June 1978, *68*, 300–10.

Uncertain Externalities, Liability Rules, and Resource Allocation: Reply

By PETER H. GREENWOOD AND CHARLES A. INGENE*

In our paper in this *Review*, we argued that the presence of uncertain externalities might cause the legal assignment of liability to affect the allocation of resources. We concluded that in an uncertain environment the Coase Theorem is valid only under a more stringent set of assumptions than are necessary for it to hold in a certain world. In their comment, Varouj Aivazian and Jeffrey Callen (hereafter A-C) raise a pair of interesting and important questions concerning the legitimacy of our assertions. First, they argue that there exists the possibility of risk sharing between the affected firms. Second, they question the potential for differential risk attitudes when transaction costs are zero. We ultimately reject the A-C implications, but only after refining our own analysis somewhat. Hence, their comments must be regarded as meritorious.

We consider only the case of legally permissible pollution; an analogous analysis holds when pollution is impermissible. Recall that in this situation the polluter, firm A, is bribed by firm B to reduce A's output. Firm A's supply function in the bribery market is $\Pi_Q = P$ and firm B's demand function is $P = (1/2)rv_Q - \bar{\pi}_Q$. ($\Pi$ and π are profits, $\bar{\pi}$ and v the mean and variance of π, Q and q are outputs, P and p are per unit bribery prices for output reduction, and R and r are risk attitudes. Subscripts denote partial derivatives. All variables referring to firm A are uppercase, those referring to B are lowercase.) The bribery market is depicted graphically in Figure 1.

I. Risk Sharing

Aivazian and Callen misinterpret the extent of our bribery market. Gains from trade are exhausted only at point E where $\Pi_Q = P^* = (1/2)rv_Q - \bar{\pi}_Q$. Gains from trade are not exhausted at the $(\hat{Q} - \tilde{Q})$ level—their $\tilde{P}_B = \$75$, $\Pi_Q = \$50 (= \tilde{P}_A)$. In bribery market equilibrium, the certain marginal profit foregone by firm A is equal to the "certainty-equivalent" value to firm B. However, A-C make the nice observation that the point E is generally Pareto suboptimal because firms A and B may trade risk in yet another private market. They overstate their case, however, by appearing to suggest that well-defined risk attitudes determine the direction of privately negotiated payments. In fact, as Coase originally argued, legal liability determines the direction of payments in the bribery market. As we demonstrate below, payment directionality in the risk-sharing market is dually determined by liability rules and the comparative levels of the absolute value of risk attitudes.

In our original Appendix B, we dealt with the private trading of risk. In light of A-C's comments, that treatment was too cursory. We now expand it. With private bribery and risk-sharing markets, firm B's profit is

$$(1) \qquad \bar{\pi}(Q,q) - P(\hat{Q}-Q) + p(\hat{q}-q) + \delta[\pi(Q,q) - \bar{\pi}(Q,q)] - b(1-\delta)$$

and A's profit is

$$(2) \qquad \Pi(Q) + P(\hat{Q}-Q) - p(\hat{q}-q) + (1-\delta)[\pi(Q,q) - \bar{\pi}(Q,q)] + b(1-\delta)$$

In (1) and (2), δ is the share of the deviation of B's profit from its expected value that B retains; $(1-\delta)$ is the share that A obtains. The per unit price of a share of B's profit deviation is b. If $b > 0$, B pays A to share the risk. If $b < 0$, A pays B for the privilege of sharing B's risk. Because A shares B's risk,

---

*University of New Hampshire and University of Texas-Dallas, respectively.

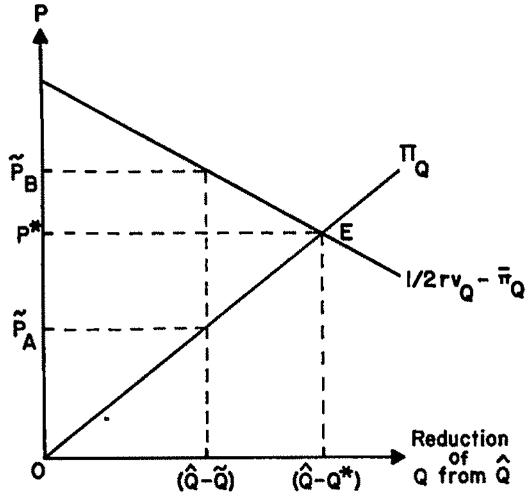


FIGURE 1

A may bribe *B* to effect a change in *B*'s output from $\hat{q}$, the unbribed level. This is represented by the $p(\hat{q}-q)$ term.

Both firms maximize their expected utility from profits. The first-order conditions with respect to the choice-theoretic variables (Q, q, δ) are given by our original (A7) and (A9). However, their right-hand sides should be multiplied by one-half as should the right-hand side of (A10). Of particular interest are the solutions for the optimal per share payment b^* and the optimal share δ^*:

$$b^* = rRv/(r+R) \tag{3}$$

$$\delta^* = R/(r+R) \tag{4}$$

Provided both firms are risk averters $(R, r>0)$, b^* is positive and δ^* lies in the positive unit interval. Firm *B* pays *A* to accept a share of its uncertain profit. When both firms are risk preferrers $(r, R<0)$, b^* is negative although δ^* lies in the positive unit interval. Firm *A* pays *B* in order to obtain a share of *B*'s profit deviation. Contrary to A-C's expectations, we see that risk is shared, not totally accepted by one firm. The direction of payment for this risk sharing is determined by legal liability, flowing from the liable to the nonliable party. However, the payment flow is reversed for dual risk preference.

When one firm is risk averse and the other a risk preferrer, we obtain a broader range of results. We first examine $r>0>R$. When *A* is, relatively speaking, a strong risk preferrer $(|R|>r)$, $\delta^*>0$, $b^*>0$. This is not substantively different from both firms being risk averters unless $|R|$ is very much greater than r. Then $\delta^*>1$, $b^*>0$. Firm *B* is paid[1] by *A* to accept more than 100 percent of its own profit deviation. In effect, *B* is being paid to gamble that the state of nature will be favorable to its actual profits π. Of course, firm *A* is betting that the state of nature will be sufficiently unfavorable to compensate it for the cost of taking such a risk-preferring position.

When $|R|<r$, $\delta^*<0$, $b^*<0$. Firm *B* is paid to accept the gamble that the state of nature will be adverse to its profits π. Firm *A* bets on the opposite position that *B*'s profits will exceed their mean value. Thus, we see that risk may be more than totally shifted and that the direction of payment is dependent upon the relative degrees of risk preference/risk aversion.

When $R>0>r$, the preceding results are partially reversed. For $R>|r|$, $\delta^*>0$, $b^*<0$. When R is very much greater than $|r|$, $\delta^*>1$, $b^*<0$. Finally, when $R<|r|$, $\delta^*<0$, $b^*>0$. The sign of b^* is uniformly reversed from the preceding case.

When firm *A* is risk neutral, $b^*=0$, $\delta^*=0$. Firm *A* accepts all the risk at no cost to *B*! When firm *B* is risk neutral, $\delta^*=1$, $b^*=0$; there is no risk-sharing market. (When $r=0=R$ there is no risk market, $\delta^*=1$.) When risk could be shared but is not (when risk is completely shifted), it must be the case that the legally liable firm (the legally nonliable firm) is risk neutral. Of course, risk neutrality guarantees that the Coase Theorem obtains unambiguously. It follows that our original Appendix B should have been incorporated into the body of our paper. With such an incorporation our qualitative results would hold, although the quantitative results would be characterized by equations such as (A10), its right-hand side divided by two.

---

[1] $-b^*(1-\delta^*)>0$.

II. Transaction Costs

We have argued that the adjusted solutions (A10) characterize the final equilibrium when firms *A* and *B* may bargain costlessly but when other parties may not enter the negotiating process. Further, the wealth levels at which *R* and *r* are evaluated differ according to the allocation of legal liability. This brings us to A-C's second point, the meaning of zero transactions costs.

When transactions are "globally" costless, there is no question that there will arise "...a 'complete' set of markets for state contingent claims" (Steinar Ekern and Robert Wilson, p. 172). In such a world, uncertainty is effectively insured away and the standard certainty results obtain. The Coase Theorem is no exception to this rule. We should have made this point clear.

When transactions are "semiglobally" costless, there may exist incomplete markets for claims. The stock market is an example. Risks are shared through the trading of securities. Managers take their risk attitude from the marketplace; indeed, all shareholders have the same marginal rate of substitution. There are a host of assumptions which must hold for the nice properties of incomplete markets to obtain. While this is not the place to delve into all the assumptions, it is worth pointing out that "[w]hen a market does not exist in which the certainty-equivalent can be established, the standard results will not in general obtain" (David Baron and Robert Forsythe, p. 573). The Arrow-Lind statements which we cited are two such situations.

An autonomous manager can exist when shareholders can gather only imperfect information about his actions, but perfect information about the ultimate outcome (the profit) of those actions. The probability of the manager retaining his position may be dependent on ultimate outcomes, not on the actions themselves. Then concern with future as well as current income may cause an autonomous manager to impose his own risk attitude upon the firm. Imperfect information may allow the manager to be successful in this endeavor.

Similarly, a dominant shareholder may possess a utility function which incorporates control as well as profits. Since other shareholders include only profits (or net income) in their indirect utility functions, there exists no market for establishing a unanimous agreement on the control/profit tradeoff. When the dominant shareholder controls over half the stock, there does not appear to be a possibility for him to be bought out by the other stockholders: the opportunity cost of "selling out" has already been discounted. Indivisibilities can block the standard results of incomplete markets.

III. Conclusion

The original Coase Theorem was devised for a world of certainty. It is based upon the twin premises of well-defined legal liabilities and "locally" zero transactions costs. Then the affected parties may bargain amongst themselves. We have again shown that these are insufficient premises under uncertainty. For the Coase Theorem to obtain in an uncertain world, it is necessary that legal liability be assigned and that there be locally zero transaction costs so that the involved parties may adjust their output levels and may share risk. But, it is also necessary that risk attitudes be independent of wealth or profit levels. This requires, at a minimum, semiglobally zero transactions costs so that a stock market may arise. In addition to the standard assumptions underlying incomplete markets, we require that there be full information with respect to managerial actions as well as the results of those actions. Alternatively, we require that shareholders be exclusively concerned with wealth. Of course, with "globally" zero transaction costs, the Coase Theorem holds unambiguously in an uncertain world.

REFERENCES

V. Aivazian and J. Callen, "Uncertain Externalities, Liability Rules, and Resource Allocation: Comment," *Amer. Econ. Rev.*, Dec. 1980, *70*, 1058–59.

D. Baron and R. Forsythe, "Models of the Firm and International Trade under Uncertainty," *Amer. Econ. Rev.*, Sept. 1979, *69*, 565–74.

S. Ekern and R. Wilson, "On the Theory of the Firm in an Economy with Incomplete Markets," *Bell J. Econ.*, Spring 1974, *5*, 171–80.

P. Greenwood and C. Ingene, "Uncertain Externalities, Liability Rules, and Resource Allocation," *Amer. Econ. Rev.*, June 1978, *68*, 300–10.

Minimum and Maximum Prices, Uncertainty, and the Theory of the Competitive Firm

By LOUIS EECKHOUDT AND PIERRE HANSEN*

Analysis of production decisions by a competitive firm under price uncertainty was published in two nearly simultaneous articles by David Baron (1970) and Agnar Sandmo. It has been extended by Baron (1971) and Hayne Leland to the case of imperfect competition. A central problem raised in these articles deals with a comparison between the output levels chosen by the firm when it is shifted from certainty to uncertainty. This has been called by some authors a "global increase" in uncertainty (see Jacques Drèze and Franco Modigliani).

The problem of a "marginal increase" in uncertainty is more difficult to handle. Sandmo, analyzing the case of perfect competition, defined this change as a mean-preserving increase in the variance of the random price, but he could not determine the sign of the reaction of output. Recently, Yasunori Ishii and Donald Coes have been able to show that nonincreasing absolute risk aversion is a sufficient condition for output to react negatively to a marginal increase in uncertainty.

In this paper, we propose an alternative definition of a marginal change in uncertainty. Its main advantage is to be more in line with real economic phenomena: the existence of guaranteed minimum and/or imposed maximum prices. The central result of the paper is obtained under the assumption of risk aversion, which is, of course, less restrictive than that of nonincreasing absolute risk aversion.

In the first section, we present the notation and discuss the effect on the output decision of minimum or maximum prices considered separately. Of course, when a minimum price is not appropriately "compensated for" by a maximum price, one cannot speak of a mean-preserving change in uncertainty, but the analysis of this question clarifies the methodology used in the following section. The second section contains the central result of the paper. We bring minimum and maximum prices together in such a way that a mean-preserving marginal increase in uncertainty is obtained and we derive its implications for production decisions. We also show that our approach leads to an alternative proof of one of Sandmo's central results. Finally, we extend the results to a situation where the expected utility for the firm (and not the expected price) is kept constant.

I. Minimum or Maximum Prices

Throughout the paper, we use Sandmo's notation: U stands for total utility and its first- and second-order derivatives are denoted by U' and U''; $\tilde{\pi}=\tilde{p}x-c(x)-B$ denotes the random profits where x is the output, $c(x)$ the total variable costs, B the fixed costs, and $\tilde{p}$ the random price; $F(p)$ is the subjective cumulative distribution of $\tilde{p}$. The density is written $f(p)$. It is non-null in the interval $[a, b]$. As usual, we assume that U' is everywhere strictly positive and U'' nonpositive. Marginal costs $(c'(x))$ are positive but may be increasing or decreasing.

If the firm attempts to maximize the expected utility of profits, its problem can be written

$$(1)\qquad \operatorname{Max} E[U]=\int_0^\infty U(px-c(x)-B)dF(p)$$

and the usual first-order condition is obtained:

---

*Professors of economics and of operations research, respectively, Institut d'Economie Scientifique et de Gestion (France) and the Faculté Universitaire Catholique de Mons (Belgium). We have benefited from useful discussion with D. Baron, M. Beuthe, A. Bultez, and P. Capéraà, and from the comments of an anonymous referee.

$$\frac{dE[U]}{dx} = \int_0^{\infty} U'\cdot(p-c'(x))dF(p) = 0 \tag{2}$$

We also assume from now on that the second-order condition is satisfied everywhere, in order to rule out local optima.

In this setting, we introduce a government agency which guarantees a minimum price for the output. If the firm thought that prices lower than the minimum one (p_m) were possible, it would revise its subjective density $f(p)$. Indeed, $f(p)$ now takes a zero value in the half-open interval $[a, p_m]$. At $p=p_m$, we associate a probability equal to $F(p_m)$, and for values of p strictly greater than p_m, the original density and the new one coincide.

After the change in density, the problem of the firm becomes

$$\max_x Z = U[p_m x - c(x) - B]F(p_m) + \int_{p_m}^{\infty} U[p\cdot x - c(x) - B]dF(p) \tag{3}$$

which gives the following first-order condition

$$\frac{dZ}{dx} = L(x, p_m) = U'_{p_m}\cdot[p_m - c'(x)]F(p_m) + \int_{p_m}^{\infty} U'\cdot[p-c'(x)]dF(p) = 0 \tag{4}$$

where the function L reminds us of the fact that dZ/dx depends on two arguments and where the subscript p_m behind U' indicates that the derivative is evaluated at the profit level corresponding to a sure price p_m.

We first observe from (4) that the optimal value of x, denoted x^*, must be such that

$$p_m < c'(x^*) \tag{5}$$

Indeed, if this were not the case, one would have $dZ/dx > 0$ and the first-order condition (4) would not be satisfied

Then, to study the sign of dx/dp_m, we notice that at the optimum $(\partial L/\partial x)dx + (\partial L/\partial p_m)dp_m = 0$. As $\partial L/\partial x$ is negative by the second-order condition, we may write

$$sgn\left(\frac{dx}{dp_m}\right) = sgn\left(\frac{\partial L}{\partial p_m}\right)$$

We thus evaluate $\partial L/\partial p_m$ and get, after obvious cancellations and division by $F(p_m)$,

$$\frac{1}{F(p_m)}\frac{\partial L}{\partial p_m} = U''_{p_m}\cdot x^*\cdot[p_m - c'(x^*)] + U'_{p_m} \tag{6}$$

If we refer to (5) and to the assumptions made about U' and U'', it is obvious that $\partial L/\partial p_m$ is positive and thus, as expected, an increase in p_m stimulates production. It is interesting to observe that the increase in output after the imposition of p_m results from the combination of two factors: an upward shift in the location of the price density and the reduction in its spread. Indeed, if the firm were risk neutral, the first term in (6) would be zero. Thus U'_{pm} stands for the positive location effect. The other term in (6) then accounts of course for the spread effect and it is positive since an increase of p_m reduces the variability of the price distribution.[1]

If we turn now to a system of maximum prices (p_M), we observe that at the optimum $c'(x^*) < p_M$ and the same kind of analysis as used above yields

$$sgn\left(\frac{dx^*}{dp_M}\right) = sgn\left(U''_{p_M}\cdot x^*\cdot[p_M - c'(x^*)] + U'_{p_M}\right) \tag{7}$$

When p_M is increased, the spread effect works against the location effect[2] and this accounts for the ambiguity of the sign of dx^*/dp_M. It is probably worth mentioning, however, that after an appropriate transformation (division by U'_{pm}), the sign of dx^*/d_{pM} is shown to depend upon the degree of "partial risk aversion" exhibited by the utility function.[3]

---

[1] We owe special thanks to the referee for pointing out this nice interpretation of (6), as well as that of equation (7).

II. A New Definition of a "Marginal Increase in Uncertainty" and its Implications

As mentioned earlier, Coes and Ishii have recently been able to show that decreasing absolute risk aversion is a sufficient condition for a "mean-preserving increase in risk" as defined by Sandmo, to reduce the output level chosen by a competitive firm. The change in uncertainty considered by Sandmo corresponds to a "stretching of the probability distribution around a constant mean." More precisely, the new random variable $\tilde{p}'$ is defined by

$$\tilde{p}'=\gamma\tilde{p}+\theta \tag{8}$$

where changes in γ and θ from $\gamma=1$ and $\theta=0$ are such that $d\theta/d\gamma=-E(\tilde{p})$ to make sure that $E(\tilde{p}')=E(\tilde{p})$.

The notion of a "stretching" of the probability distribution has some intuitive appeal, but it is hard to think of exogenous factors which would yield this specific change in the subjective expectations of the firm. This is the reason we prefer to approach changes in risk in terms of minimum and maximum prices which are institutional factors modifying the price expectations of the firm. Of course, we agree that many price support programs are financed by contributions of the firm in various forms other than maximum prices. However, there are also examples of a combination of minimum and maximum prices such as the limits for exchange rate fluctuations, or agricultural agreements for some types of products at the European Economic Community level. Thus it may be more realistic to approach changes in risk in such terms. Besides, as it will become clear later, more general results can be obtained from our definition.

A mean-preserving "price squeeze" through minimum and maximum prices can be defined as follows. Let the original density function with expectation $E(\tilde{p})$ take nonzero values in the interval $[a,b]$. It is then possible to find a minimum price $p_m(p_m>a)$ and a maximum one $p_M(p_M<b)$ such that $E(\tilde{p})$ remains constant. More formally, p_m and p_M are given by the solution of the following equation:

$$\int_a^b pf(p)dp=F(p_m)p_m+\int_{p_m}^{p_M}pf(p)dp+(1-F(p_M))p_M \tag{9}$$

This definition turns out to be a special case of another definition proposed by Michael Rothschild and Joseph Stiglitz. Indeed, equation (9) meets the requirements of the "integral condition" they proposed. Our equation (9) also leads to the notion of a global change in uncertainty which is obtained by choosing p_m and p_M as equal to $E(\tilde{p})$.

Notice also that, to discuss "marginal changes," we have to differentiate (9) totally with respect to p_m and p_M and, in order to keep $E(\tilde{p})$ constant, we get

$$\frac{dp_M}{dp_m}=-\frac{F(p_m)}{1-F(p_M)} \tag{10}$$

---

[2] Indeed, it can be shown that an increase in p_M increases the variance of $\tilde{p}$.

[3] Indeed, partial risk aversion (P) is defined by

$$P=-t\frac{U''(W+t)}{U'(W+t)}$$

where t and W have the dimension of money. If one writes

$$W+t=p_Mx^*-c(x^*)-B$$

and

$$t=p_Mx^*-x^*c'(x^*)$$

and, if one divides the right-hand side of (7) by U'_{pM}, it is easily seen that a decrease in p_M stimulates output when P is evaluated at t, and $W+t$ is greater than unity in absolute value.

When the firm faces a system of minimum and maximum prices, its objective function is written

$$(11)\quad \underset{x}{\operatorname{Max}} H = U[p_m x - c(x) - B]F(p_m) + \int_{p_m}^{p_M} U[px - c(x) - B]\,dF(p) + U[p_M x - c(x) - B]\cdot[1 - F(p_M)]$$

and the first-order condition becomes

$$(12)\quad \frac{dH}{dx} = L(x, p_m, p_M) = U'_{p_m}\cdot[p_m - c'(x)]F(p_m) + \int_{p_m}^{p_M} U'\cdot[p - c'(x)]\,dF(p) + U'_{p_M}\cdot[p_M - c'(x)][1 - F(p_M)] = 0$$

Inspection of (12) reveals that at the optimum

$$(13)\quad p_m < c'(x^*) < p_M$$

Considering now a change in p_m and p_M with $E(\tilde{p})$ constant, we have

$$(14)\quad \frac{\partial L}{\partial x}dx + \frac{\partial L}{\partial p_m}dp_m + \frac{\partial L}{\partial p_M}\cdot\frac{dp_M}{dp_m}dp_m = 0$$

and thus, given the negative value of $\partial L/\partial x$ and because of (10),

$$(15)\quad sgn\left(\frac{dx^*}{dp_m}\right)\bigg|_{dE(\tilde{p})=0} = sgn\left(\frac{\partial L}{\partial p_m} - \frac{\partial L}{\partial p_M}\cdot\frac{F(p_m)}{1 - F(p_M)}\right)\bigg|_{dE(\tilde{p})=0}$$

where $dE(\tilde{p}) = 0$ reminds us of the fact that p_m and p_M are modified simultaneously in accordance with (10).

Evaluating $\frac{\partial L}{\partial p_m}$ and $\frac{\partial L}{\partial p_M}$, we finally get

$$(16)\quad \frac{\partial L}{\partial p_m} - \frac{\partial L}{\partial p_M}\cdot\frac{F(p_m)}{1 - F(p_M)} = U''_{p_m}\cdot x^*[p_m - c'(x^*)]F(p_m) - U''_{p_M}\cdot x^*[p_M - c'(x)]F(p_m) + U'_{p_m}F(p_m) - U'_{p_M}F(p_m)$$

The first two terms in this expression are positive because of (13). Besides, the concavity of U implies that U'_{pm} exceeds U'_{pM} so that the difference between the last two terms is also positive. We thus obtain the central result of the paper.

PROPOSITION 1: *A mean-preserving price squeeze through minimum and maximum prices as defined in* (10) *leads to an increase in output for a risk-averse firm.*

It is worth making a comparison between this result and those obtained by Sandmo, Coes, and Ishii with the help of their definition of a mean-preserving increase in variance. In fact, we arrive at the same kind of conclusion, but on the basis of a much less restrictive assumption: with a price squeeze, *risk aversion* is a sufficient condition for a decrease in risk to stimulate production, while in Sandmo's approach, *decreasing risk aversion* is a sufficient condition to obtain the same result.

Our analysis can be viewed as an alternative proof of one of Sandmo's main results. Indeed, he has shown that the competitive firm would produce more if it faced a sure price equal to $E(\tilde{p})$ than in the presence of a random price with expectation $E(\tilde{p})$. In our model, if we continuously change p_m and p_M, we will finally arrive at a sure price equal to $E(\tilde{p})$, and because of Proposition 1, the output corresponding to this sure price will be higher than any output level chosen before. Thus Proposition 1 contains, at the limit, Sandmo's result about a global change in uncertainty.

The interest of defining changes in risk in terms of price squeeze also appears from another consideration. Peter Diamond and Stiglitz have argued convincingly in favor of the analysis of what they call "a mean-utility-preserving increase in risk" (rather than the usual mean-preserving one). Let us now consider in our framework a simultaneous change in p_m and p_M such that the optimal level of expected utility remains constant. The analysis is very much like the one developed above, except that the ratio dp_M/dp_m takes a different value.

To determine the values of dp_m and dp_M which leave $E[U]$ constant, we totally differentiate the function H in (11) with respect to p_m and p_M and, if we keep in mind that dH/dx equals zero since we are at the optimum, we get

$$U'_{p_m}\cdot x^*F(p_m)dp_m + U'_{p_M}\cdot x^*[1-F(p_M)]\,dp_M=0 \tag{17}$$

or

$$\frac{dp_M}{dp_m}=-\frac{U'_{p_m}\cdot F(p_m)}{U'_{p_M}\cdot[1-F(p_M)]} \tag{18}$$

As could be expected, (18) is very similar to (10), except that the probabilities are weighted by the marginal utilities of the profit level in order to keep $E[U]$ rather than $E(\tilde{p})$ constant.

The introduction of (18) into equation (14) yields, after the same simplifications as before, the following proposition:

PROPOSITION 2: *A mean-utility-preserving decrease in uncertainty (increase in p_m and decrease in p_M) leads to an increase in output for a risk-averse firm.*

Again, this result is obtained under a plausible assumption: the existence of risk aversion. This contrasts with the fact that Diamond and Stiglitz have to place rather severe restrictions on the shape of U to get the same result. However, it must be recalled that our price squeeze is a special case of their "integral condition" which enables them to rank more distributions than does our approach.

III. Concluding Remarks

We have proposed in this paper an alternative definition of a marginal increase in uncertainty, and we have applied it to the production decisions of a competitive firm. This definition calls upon the notions of minimum and maximum prices which we have combined to define a new type of "mean-preserving" decrease in uncertainty. Our new definition has led us to the central proposition of the paper according to which a decrease in uncertainty stimulates the output of a risk-averse firm. The same result holds if the decrease in uncertainty is such that the expected utility of the firm (and not its expected price) remains constant. It thus appears that our definition of a marginal change in uncertainty is rather "robust," at least for this kind or problem.

REFERENCES

D. Baron, "Price Uncertainty, Utility, and Industry Equilibrium in Pure Competition," *Int. Econ. Rev.*, Oct. 1970, *11*, 463–80.

______, "Demand Uncertainty in Imperfect Competition," *Int. Econ. Rev.*, June 1971, *12*, 196–208.

D. Coes, "Firm Output and Change in Uncertainty," *Amer. Econ. Rev.*, Mar. 1977, *67*, 249–51.

P. Diamond and J. Stiglitz, "Increases in Risk and in Risk Aversion," *J. Econ. Theory*, July 1974, *8*, 337–60.

J. Drèze and F. Modigliani, "Consumption Decisions under Uncertainty," *J. Econ. Theory*, Dec. 1972, *5*, 308–35.

Y. Ishii, "On the Theory of the Competitive Firm under Price Uncertainty: Note," *Amer. Econ. Rev.*, Sept. 1977, *67*, 768–69.

H. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

M. Rothchsild and J. Stiglitz, "Increasing Risk: I. A Definition," *J. Econ. Theory*, Sept. 1970, *2*, 225–43.

A. Sandmo, "On the Theory of the Competitive Firm under Price Uncertainty," *Amer. Econ. Rev.*, Mar. 1971, *61*, 65–73.

Estimating Demand Elasticities for Public Goods from Survey Data

By BETTY BLECHA GIBSON*

In earlier issues of this *Review*, Thomas Borcherding and Robert Deacon, and Theodore Bergstrom and Robert Goodman, use similar models derived from collective choice theory to estimate demand elasticities for public goods. The present note also estimates demand elasticities for public goods, in this case for six categories of expenditures by the State of California. But the supporting model differs from the earlier models in important respects:

1. The model uses expressed preferences made by consumers in different tax, price and income situations in a given polity, not the choices of median voters in different polities.
2. A taste variable is included in the model, and the heretofore speculative significance of taste differences in demand functions for public goods is evaluated.
3. Problems of observing individual preferences for public goods are solved by imposing restrictive assumptions on a survey research design.

The estimated elasticities have appropriate signs and reasonable values. Of particular interest are the findings of a significant taste variable, and of a low value for the income elasticity for "aid to needy people." The results permit me to extend and clarify several aspects of the two previous studies. For example, the estimates are shown to support the contention that Borcherding and Deacon's positive price elasticities for highways and higher education are due to an invalid tax price measure. Finally, the results show that survey research can be used to test public goods demand models that require an interval level of measurement. The model itself thus provides an important methodological extension in an area where research has been handicapped by the lack of observable interval information on individual consumers.

I. The Model

To develop the model, it is necessary to specify each consumer i's demand function for a state-produced activity z. The following assumptions are used to derive each consumer i's demand function: Each consumer i pays the fraction t_i of the total cost of state expenditures. Each t_i is invariant with respect to the size of state expenditures. Activity z is supplied at constant marginal cost q_z. Each consumer i knows his tax price for z, $t_i q_z$, and computes z_i, his preferred quantity of z, by maximizing his utility subject to his budget constraint and the given tax price. Each consumer's taste index for z is w_i. The taste index w is defined such that if two consumers have the same incomes and confront the same prices, but differ in terms of w, the consumer with the higher w prefers the larger quantity of z. While the definition of w is consistent with the general usage of taste differences, many functional forms of w in each consumer's utility function are consistent with the definition. Because it will not be possible to determine the precise form from the data set, w is assumed to enter each consumer's demand function in the form $c(w_i)$, where $c(w)$ is a monotonic increasing function of w. Let there be constant price and income elasticities α and β for activity z, and let y_i be each consumer i's income. Then each consumer i's demand function for z is

$$z_i = ac(w_i)(t_i q_z)^{\alpha} y_i^{\beta} \tag{1}$$

---

*Assistant professor of economics, University of San Francisco. The research was partially supported by a grant from the San Jose State University Foundation. I wish to thank Thomas Pogue for suggestions on an earlier version of this paper. I also wish to thank the referee and managing editor for helpful comments.

This demand function determines each consumer i's preferred quantity of z, given his tax price, income, and taste index.

It is in general not possible to observe each consumer i's preferred quantity of z because of the publicly supplied nature of z. The actual quantity of z supplied to the community is determined collectively, and each consumer adjusts his consumption to what he perceives to be the collectively supplied quantity regardless of whether or not this is his preferred quantity. In the Borcherding-Deacon and Bergstrom-Goodman studies, the problem of observing preferred quantities is solved by making a series of collective choice assumptions. These assumptions insure that the quantity of a publicly supplied activity actually chosen by the community is determined by the median voter for the community. It is further assumed that the median voter receives the median income.[1] When combined with the assumptions of constant t_i and q_z, this median income assumption allows the expenditures of the community for z to be treated as an observation on the demand function of the median income consumer. The demand function is assumed to be *log-linear* in both studies. With the exception of the taste index, equation (1) follows the demand function used by Borcherding-Deacon and Bergstrom-Goodman.[2]

By assuming that median income consumers in different communities are on their demand curves, Borcherding-Deacon and Bergstrom-Goodman were able to estimate demand parameters for median voters in municipalities and states. In contrast, my study identifies within a sample of consumers in a single jurisdiction a subset of consumers *who perceive themselves to be consuming on their demand function for a publicly supplied activity*. Demand parameters are then estimated from observations on this subset of consumers. A survey research instrument is used to identify the subset and measure the necessary variables.

The assumptions needed to utilize a survey research design are as follows: each consumer i who perceives himself to be in equilibrium with respect to z is asked to indicate the benefit b_i he derives from consuming z. He is given a benefit scale and asked to indicate the value that best describes his benefit b_i. The benefit scale is restricted to vary between two predetermined real numbers. It is assumed that each consumer i responds by first computing an income equivalent of the utility he derives from consuming z. It is further assumed that he uses the area under his ordinary demand curve for z to estimate his income equivalent. The shaded area in Figure 1 illustrates each consumer i's estimate of his income equivalent, where z_i^* is consumer i's perception of the collectively determined quantity of z. It is also assumed that each consumer i states his income equivalent relative to his income in responding to the benefit question. Finally, it is assumed that each consumer i uses the proportional rule k to map his ratio of income equivalent to income into the predetermined range of responses for b_i. Let o_i be consumer i's estimate of his income equivalent. Then each consumer i's benefit response is

$$b_i = k(o_i/y_i) \tag{2}$$

Solving for o_i by integration,[3] substituting for z from (1) to eliminate z, and simplifying

---

[1] See the two studies for a complete discussion of the various assumptions.

[2] The constant elasticity form is well established in public goods investigations. The specification used here does not include a variable denoting crowding (or publicness). These variables were included in the Borcherding-Deacon and Bergstrom-Goodman studies, where they were measured by population. Since the data examined here are drawn from a single jurisdiction, no comparable variable can be defined.

[3] The integrand for the demand function does not have a finite value for all values of α. But it seems reasonable to assume that consumers will give a finite estimate of their income equivalent when the integrand is not finite. In these cases, it is assumed that each consumer i estimates his income equivalent o_i by some multiple of the amount he pays for z_i^*. Then a statement is generated that differs from (3) only in the constant term. Each consumer i pays $(t_i q_z) z_i^*$ for z_i^*. Let d be the multiple term equating what each consumer i pays for z_i^* to o_i. Then $o_i = d(t_i q_z) z_i^*$. Substituting for z_i^* from (1) yields $o_i = d' c(w_i) t_i^{\alpha+1} y_i^{\beta}$. But $b_i = k(o_i/y_i)$ and $b_i = d'' c(w_i) t_i^{\alpha+1} y_i^{\beta-1}$.

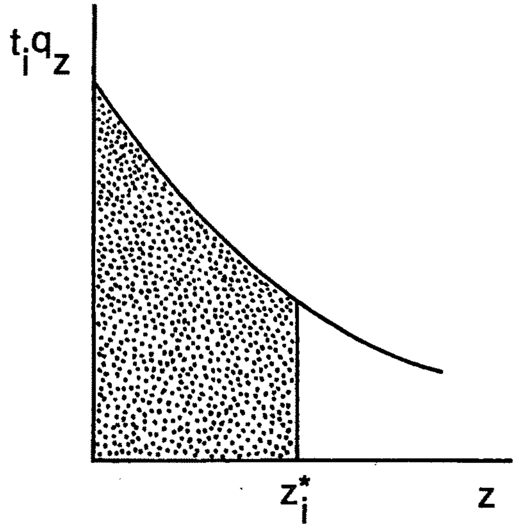


FIGURE 1

gives

$$b_i = k'c(w_i)t_i^{\alpha+1}y_i^{\beta-1} \tag{3}$$

Given information on b, w, t, and y for consumers in equilibrium with respect to a public good, it is possible to estimate price and income elasticities from the logarithmic form of (3) with the addition of a random disturbance term.

The behavioral assumptions used to derive (3) are admittedly assumptions of convenience that allow the model to be estimated by a survey research design. But the behavioral assumptions are not unrealistic. When asked to evaluate the benefit derived from a good, it is reasonable to expect the consumer to think in terms of indicating his utility by some dollar amount. Further, it is reasonable to expect this income equivalent to be stated in some relative way. How much the consumer benefits from consuming z is relative to the benefit he derives from all other goods, which is assumed here to be measured by income.[4]

Assuming the consumer computes an income equivalent of his utility from consuming z, economic theory does not provide a unique answer on what the income equivalent will be. The computation depends on the value of the marginal utility of income along the adjustment path between the consumer's initial and final position, with the Hicksian expressions of compensating and equivalent variation generally providing the upper- and lower-limiting values.[5] Given that the consumer does not know what to assume for an adjustment path when asked a benefit question, it is reasonable to argue that he will adopt some path or construct between the two limiting values. The obvious construct is the consumer's ordinary demand curve. Finally, when asked to map his ratio of income equivalent to income into some predetermined range of magnitudes, it is reasonable to expect the consumer to choose a rule that leaves his responses unaffected by any linear change in magnitude assignment. A proportional rule has this desired property. It is assumed for convenience that each consumer uses the same proportional rule.

II. Measurement of Variables

Observations needed to estimate (3) are obtained from a systematic sample of registered voters in Santa Clara County, California. A survey was administered to the sample during the two-week period preceding November 6, 1973. The return rate was 60 percent, with 182 respondents being interviewed. Only one variable, party identification, was available to check for representativeness. At the time the sample was drawn, 38 percent of the registered voters were Republican, 55 percent were Democrats, and 7 percent showed some other affiliation. The corresponding values for the sample were 38, 58, and 5 percent, respectively. The interviewing was done mainly by San Jose State University students. Each interviewer had two training sessions before going into the field.

To estimate (3), it is first necessary to identify the subsets of respondents who are on their demand curves with respect to each of six state activities. This is accomplished by asking respondents whether they favor increasing, decreasing, or leaving unchanged state spending for each activity if the spending change is reflected in their state tax payments.[6] By linking a change in expenditures to a change in subjective tax price, it is possible to infer whether a respondent's subjective tax price is less than, greater than, or equal to his perceived marginal benefit. A respondent whose tax price is equal to his perceived marginal benefit is consuming on his demand curve for the activity and presumably answers no change. Only respondents who answer no change are used to estimate (3) for the activity, with there being different subsets of respondents for each activity.

---

[4] It could also be assumed that the benefit the consumer derives from all other goods is proportional to income. The resulting equation would differ from (3) only in the constant term.

[5] See Eugene Silberberg and Michael Burns.

The variables needed to estimate (3) for the subsets of respondents are measured as follows: to measure benefits derived from an activity, respondents are asked to rank the benefits they feel they derive from each of the six activities on a seven-point scale.[7] Income is measured by using the midpoint of the income category respondents indicate includes their family income for the current year.[8] Perceived tax share is measured by perceived tax liability. As long as the marginal cost of each state activity is constant, this substitution of tax liability for tax share does not affect elasticity estimates, although the constant term is affected.[9] Respondents are asked how much they feel they will pay in state sales and income taxes during the current year. These two values are then added together to form the perceived tax liability variable.[10] While other revenue sources exist, it is not possible to assign the incidence of these other sources to respondents in a convincing way. But the high percentage of general fund expenditures financed by sales and personal income taxes (69 percent for the study period) justifies the tax liability measure. The measure also has the strength of directly reflecting perceived tax liability. Borcherding-Deacon and Bergstrom-Goodman estimated perceived tax shares from aggregate data, and the possibility that nonrandom perceptual errors were causing bias could not be dismissed.

A Likert-type attitude scale is used to measure the taste index.[11] The scale items are shown in Table 1. Each item has five possible responses running in an index from 1 to 5, with 5 representing the strongest taste. Each respondent is given a summed score based on his responses to the items. For example, a respondent answering "a great deal" to items 1, 2, and 4 and "very little" to items 3 and 5 receives a score of 25. This score is used to measure $c(w_i)$.

---

[6] The survey question was worded: "I'm now going to ask you if you think state spending should increase, decrease, or stay about the same for several state programs. People often wonder where the money will come from if spending is increased for some programs, or what will happen to savings from cutting back spending on some programs. In answering these questions we would like you to think that an increase in spending for a program would mean some increase in your taxes, and a decrease in spending for some program would mean some decrease in your taxes. In other words, if you think state spending on pollution control should increase, this increase would mean some increase in your state taxes. If you think state spending on pollution control should decrease, this decrease would mean some decrease in your state taxes. I'll read the programs off one at a time and you just tell me if you think spending should increase, decrease, or stay about the same."

[7] The survey question was worded: "I would like to begin by asking you how you feel about certain things the State of California spends money on. (HAND R CARD #1.) This card is like a thermometer. Someone who felt he benefited a great deal from some state program like highway construction would place himself at seven. Someone who felt he didn't benefit at all from highway construction would place himself at one. Of course other people have feelings in between. I'm going to read off a list of state programs one at a time. You just tell me which number best describes your feelings toward this program or service."

[8] The income question is a standard Survey Research Center item with the following income groups: under $1,000; $1,000–1,999; $2,000–2,999; $3,000–3,999; $4,000–4,999; $5,000–7,499; $7,500–9,999; $10,000–14,999; $15,000–24,999; $25,000 and over. The midpoint for the last category was assumed to be $30,000.

[9] Let e be the total cost of state expenditures and let x_i be consumer i's perceived tax liability. Then $t_i = x_i/e$ and $\ln(x_i/e)^{\alpha+1} = (\alpha+1)\ln x_i - (\alpha+1)\ln e$. But $(\alpha+1)\ln e$ is a constant.

[10] The survey question was worded: "We are interested in what people think they pay in different taxes over a year's time. I'm going to read off several different taxes one at a time and you just tell me how much you think you will pay in this tax over this year. Just give me your best guess."

[11] Complete discussions of scaling are found in Warren Torgerson and Clyde Coombs. See H. Upshaw for a brief discussion.

TABLE 1—TASTE SCALE

Item		Direction of Strong Taste
1.	How much do you think state spending on education benefits those families without children in school?	a great deal
2.	Some people feel they benefit indirectly from a great many things state government does even though these programs don't help them directly very much. How much do you feel you benefit indirectly like this from things state government does?	a great deal
3.	Some think that the only people who really benefit from state parks are people who use them. How much truth do you think there is to this way of thinking?	very little
4.	How willing are you to pay taxes to support state programs which don't benefit you or your family directly but which do benefit other people in the state?	a great deal
5.	People sometimes say they only benefit from state government services when they use these services. How much truth do you think there is to the way these people feel?	very little

Note: The possible responses were: a great deal, pretty much, some, little, and very little.

However, two questions must be answered in the affirmative before the summed score can be used to estimate the model: Do the five items form a scale? If so, is the researcher justified in treating the scale as a taste measure?

Because there is no single, generally accepted test for the scalability of Likert items, several different tests are applied to the five items. All tests support the assertion that the items do form a scale. Interitem to scale correlation coefficients and a *t*-test of the scale's ability to discriminate the upper 25 percent of the scores from the lower 25 percent are significant at the .05 level. Two different kinds of factor analysis also support the scalability of the items. A factor analysis of the items with squared multiple correlation coefficients used as estimates of communalities generates one factor with loadings for items 1 through 5 of .666, .698, −.367, .620, and −.518, respectively.[12] A principal-component analysis generates two factors that account for 68 percent of the total variance. The second factor loads heavily on items 3 and 5 and has an eigenvalue of 1.03. It is concluded that some negative response bias is present but that the large portion of variance accounted for by the first factor supports the argument that the scale is measuring a single variable. It should be noted that a unique interval meaning cannot be given to an estimated coefficient of the *log* of the taste measure. The interval properties of the scale are suspect, with no more than an ordinal level of measurement usually assumed for Likert scales. Moreover, any monotonic increasing function of w is consistent with the definition of the taste index in the model. Thus the interval properties assumed for $c(w_i)$ in order to justify the *log*-linear form represent a convenient, but perhaps not very realistic, assumption.

Given that the five items form a scale, is it reasonable to suggest that the scale measures tastes? It has been assumed that consumers interpret benefit questions in terms of an income equivalent. The taste index has been defined such that a positive relationship exists between the taste index and the preferred quantity of a state activity, all other things being equal. It follows that a positive relationship exists between the taste index and the income equivalent associated with a given state activity. This variation in income equivalent will then be reflected in different benefit responses. The scale items directly reflect the benefit variation expected to follow from taste differences and thus can be used to measure the taste index.[13]

---

[12] The categories for items 3 and 5 were not recoded to reflect the direction of strong taste consistent with the categories for items 1, 2, and 4 when the factor analyses were performed.

[13] A formal discussion of the scale design is lengthy and available from me on request. Briefly, the taste index examined in the text assumes that state-produced activities supply two types of goods in fixed propor-

TABLE 2—REGRESSION RESULTS

	Taste Variable Excluded				Taste Variable Included					
State Activity	Constant	Price Elasticity	Income Elasticity	R	Constant	Price Elasticity	Income Elasticity	Taste Coefficient	R	Number of Cases
Aid to Needy People	7.33	−.78 (.14)	.21 (.27)	.41	5.89	−.83 (.15)	.26 (.27)	.45 (.35)	.44	47
Pollution Control	3.04	−.88 (.19)	.74 (.36)	.13	1.17	−.99 (.19)	.77 (.34)	.84 (.43)	.37	32
Colleges and Universities	1.46	−.82 (.12)	.88 (.21)	.22	−1.19	−.87 (.11)	.92 (.19)	.96 (.28)	.49	51
Elementary School Aid	1.39	−1.07 (.13)	1.04 (.21)	.09	−1.29	−1.16 (.12)	1.14 (.20)	.87 (.28)	.41	52
Parks and Recreational Areas	.90	−1.03 (.10)	1.08 (.19)	.07	.20	−1.02 (.10)	1.06 (.18)	.32 (.21)	.22	50
Highway Construction and Maintenance	2.15	−1.08 (.08)	1.00 (.06)	.20	1.02	−1.09 (.08)	.99 (.06)	.42 (.29)	.32	33

Note: Standard errors are shown in parentheses. Elasticity estimates and their standard errors are derived from the regression estimates and standard errors for $\alpha+1$ and $\beta-1$.

The measures for income, benefit, and taste are all categorized measures. The categories are defined by the scale intervals, and the category midpoints are assumed to equal the means of the cases distributed within the intervals. For example, if a respondent indicates a benefit score of 5, his actual score is assumed to lie in the interval 4.5 to 5.5, and 5 is assumed to be the mean benefit score for all respondents in the interval. Thus the observed measures for the categorized variables reflect measurement error. While measurement error in the benefit variable can be assumed to be captured in the random disturbance term, measurement error in the independent variables of income and taste will cause ordinary least squares (*OLS*) estimators to be biased and inconsistent. In order to assess the effect of measurement error, a maximum likelihood method is applied to solve for the estimates of (3).[14] Estimates of the maximum possible measurement error present in each regression are obtained by assuming that all cases lie at the category bounds. The resulting squared deviations about the category midpoints are weighted by the proportion of the total regression cases occurring in the categories to obtain estimates of error variances for the income and taste variables. These estimates are so small in relation to the variances for the income and taste variables that the maximum likelihood method generates essentially the same estimates as *OLS*.

III. Regression Results

Equation (3) is estimated for each of the six state activities. The regression results are shown in Table 2, with the benefit measures for each activity being the dependent variables. A large amount of missing data for the tax variable and the estimation of (3) for only those respondents on their demand curves combine to reduce the number of cases from the initial 182 respondents for all six regressions.

All estimates shown in Table 2 have appropriate signs and reasonable values. Given

---

tions. The two types of goods are defined as follows: if consumers can derive benefits from a good supplied by an activity only by using the activity, the good is a direct consumption good; if consumers do not have to use the activity to derive benefits from a good, the good is an indirect consumption good. The scale shown in Table 1 measures tastes for only indirect consumption goods. But it is easily shown that if consumers have the same utility function the preferred level of an activity varies directly with the taste index for indirect consumption goods. Thus the taste index can be included in a *log*-linear demand function along with tax price and income, as shown in (1).

[14] See Jack Johnston, pp. 287–90.

the use of categorized survey data, multiple *R*s are predictably low. But the failure of the tax liability and income variables to explain a larger amount of the variance for parks and recreational areas and elementary school aid is surprising. It should be noted that regressions for parks and recreational areas have not behaved consistently in both the Borcherding-Deacon and Bergstrom-Goodman studies. The regressions for parks and recreational areas had a low multiple *R* for one of Borcherding-Deacon's groups, and Bergstrom-Goodman found only one significant tax share elasticity in their ten state regressions, along with a nonsignificant low-income elasticity for California. These combined results lead to the speculation that significant variables are being excluded in demand specifications for parks and recreational areas. Variables such as travel distance to publicly financed parks, congestion costs, availability of private substitutes, and availability of free substitutes would be expected to have a substantial impact on demand. In the case of elementary school aid, substantial local financing and quality differences among school districts within Santa Clara County could explain the failure of the tax liability and income variables to account for more variance. Respondents may not be able to clearly distinguish state aid shares from local shares or relate state aid to elementary school output. Quality differences among school districts would imply that state aid does not constitute the same good among districts.

Some of the elasticities shown in Table 2 relate to specific results found in previous studies. Borcherding and Deacon reported positive price elasticities for higher education and highways. They speculated that their results for these highly subsidized activities were due to the lack of association between marginal cost (used to estimate tax share) and the median voter's perceived tax share. The negative price elasticities for universities and highways shown in Table 2 are based directly on perceived tax liabilities and thus support Borcherding and Deacon's reasoning. Also, to the best of my knowledge, the elasticities for aid to needy people represent the first estimates for a category of welfare expenditures to be based on individual consumer data. Previous estimates have been based on the actual output of the polity and show substantially higher income elasticities.[15] While conclusions drawn from the estimates reported in Table 2 must be tenuous, the apparent discrepancy between consumer data and the output of the polity poses a significant question for future research. Finally, the regression results support the speculation that taste differences have a substantial impact on demand for public goods. Indeed, a conservative conclusion is that the taste variable is at least as important as tax liability and income in five of the six regressions.[16]

IV. Implications

The regression results have implications for both specifications of demand information in collective choice studies and future applications of survey-based methodology. The implications for demand specifications are discussed first. In evaluating constant elasticity functions for public goods, the sign of $\beta+\alpha\eta$ has important implications, where η is the elasticity of tax share with respect to income.[17] If $\beta+\alpha\eta>0$ (<0), quantity demanded is an increasing (decreasing) function of income, a necessary condition for many median voter applications to collective choice problems. Bergstrom and Goodman found β to be substantially larger than $\alpha\eta$ for most of their regressions (assuming a tax share elasticity in the range of 1.0 to 1.3). Applying the same tax share elasticities to Borcherding and Deacon's estimates shows only half of their regressions yield a β substantially greater than $\alpha\eta$. Given the standard errors, caution must be exercised in drawing inferences for activities shown in Table 2. But $\beta+\alpha\eta$ for aid to needy people is clearly negative for any reasonable assumption about state tax elasticity, and the combined results suggest that it may be premature to speak generally of the demand for public goods or expect consumers with incomes above (below) the median to vote for increasing (decreasing) public expenditures.

---

[15] See Larry Orr for a discussion of estimating income elasticities and income effects for Aid to Families with Dependent Children.

[16] The regression results for aid to needy people are surprising only if it is argued that a redistributive taste should be reflected in the taste index. To examine this argument, it was assumed that redistributive tastes are associated with preferences concerning whether a progressive or proportional tax system is more equitable. A question on tax equity was included on the questionnaire and associated with the scale. No systematic association between this question and the scale existed. Thus the regression results for aid to needy people are consistent with other redistributive relationships in the data set.

[17] See Bergstrom and Goodman.

The exclusion of taste variables in previous specifications raises questions of possible bias in previous estimates. No bias problems exist as long as the taste variables are uncorrelated with other independent variables. The taste variable used in this paper is uncorrelated with tax liability and income in the data set.[18] This result gives added support to the Borcherding-Deacon and Bergstrom-Goodman estimates. But the relationship may not be typical of other data sets. Santa Clara County has one of the wealthiest best-educated populations in the United States. Normal survey bias and use of the voting roles as the operational population reinforced these characteristics in the sample. Thus the sample shows little evidence of socioeconomic cleavages. In polities where such cleavages are substantial, the taste variable may be correlated with income. Even for the Santa Clara County sample, a careful examination of the data set shows a significant ordinal association between the high-income category and other income categories with respect to the taste scale.[19]

The regression results also have implications for the survey-based methodology. The reasonable estimates provide some support for the assumptions used to generate the model. If the theoretical structure of the model is accepted, any public choice hypothesis derived from demand theory is potentially testable with survey research. The problem of not being able to observe individual choices, long a handicap to research in collective choice behavior, is removed. The model does not, however, remove problems due to strategic behavior. Consumers are still assumed to reveal their true preferences, and care must be taken not to apply the model to those cases where consumers are cognizant of strategic gain.

REFERENCES

T. Bergstrom and R. Goodman, "Private Demands for Public Goods," *Amer. Econ. Rev.*, June 1973, *63*, 280–96.

T. Borcherding and R. Deacon, "The Demand for the Services of Non-Federal Governments," *Amer. Econ. Rev.*, Dec. 1972, *62*, 891–901.

M. Burns, "A Note on the Concept and Measure of Consumer's Surplus," *Amer. Econ. Rev.*, June 1973, *63*, 335–44.

Clyde H. Coombs, *A Theory of Data*, New York 1964.

L. A. Goodman and W. A. Kruskel, "Measures of Association for Cross Classification III: Approximate Sampling Theory," *J. Amer. Statist. Assn.*, June 1963, *58*, 310–64.

Jack Johnston, *Econometric Methods*, 2d ed., New York 1972.

L. Orr, "Income Transfers as a Public Good: An Application to AFDC," *Amer. Econ. Rev.*, June 1976, *66*, 359–71.

E. Silberberg, "Duality and the Many Consumer's Surpluses," *Amer. Econ. Rev.*, Dec. 1972, *62*, 942–52.

Warren S. Torgerson, *Theory and Methods of Scaling*, New York 1958.

H. Upshaw, "Attitude Measurement," in Hubert M. Blalock, Jr. and Ann B. Blalock, eds., *Methodology in Social Research*, New York 1968.

---

[18] Correlations for the taste variable and income are: aid to needy people .00; pollution control .29; colleges and universities .04; elementary school aid .03; parks and recreational areas .06; highway construction and maintenance .23. Correlations for the taste variable and tax liability are: aid to needy people .15; pollution control .41; colleges and universities .13; elementary school aid .19; parks and recreational areas .01; highway construction and maintenance .21.

[19] The lower-income groups were collapsed into one, and Goodman and Kruskal's *gamma* was computed. *Gamma* was .24 and significant at the .05 level.

The Value of a Life: New Evidence from the Marketplace

By RACHEL DARDIS*

Problems of life valuation have been discussed extensively in recent years (see Jan Acton; Brian C. Conley; Jack Hirshleifer, Theodore Bergstrom, and Edward Rappaport; M. W. Jones-Lee, J. Linnerooth, E. J. Mishan, Thomas Schelling, Richard Thaler and Sherwin Rosen, Dan Usher, and Richard Zeckhauser). Various approaches have been discussed, including the individual's willingness to pay for a risk reduction. This approach is favored by many economists on the grounds that it incorporates the preferences of the individual towards risk.

Procedures for estimating the value of risk reduction to the individual include: 1) questionnaires; 2) risk premiums for workers in hazardous occupations; and 3) voluntary purchases of safety devices or products with safety features. While the questionnaire method and the risk premium method have been employed by several researchers, lack of market data has precluded the use of the third method. This paper investigates voluntary purchases of smoke detectors to estimate consumer willingness to pay for risk reduction. Such data may in turn be used to estimate the implicit value of a life.

I. Willingness-To-Pay Model

The willingness-to-pay measure of life valuation is shown in Figure 1. Utility is measured along the vertical axis while income is measured along the horizontal axis. In the initial risk-free situation, the individual is at point A where consumption C_0 yields a utility of $U(C_0)$. The introduction of a hazard replaces the certain prospect C_0 with an uncertain prospect which yields income C_0 with a probability $(1-p)$, and zero income or death with a probability p. Assuming that the state "death" yields zero utility in the *ex ante* calculations of the decision maker (i.e., there is no bequest utility function), then the uncertain prospect has an expected utility (EU) equal to $p(0)+(1-p)UC_0$.[1] The loss in utility is equal to $pU(C_0)$. This loss in utility may be translated into monetary terms by considering the amount of wealth necessary to compensate the individual for the introduction of the hazard. This is given by C_0-C_1 in Figure 1.

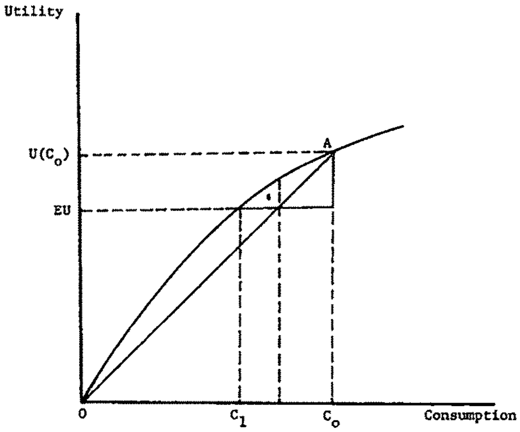


FIGURE 1. DERIVATION OF THE IMPLICIT VALUE OF A LIFE

For small changes in the neighborhood of A, the compensating variation $V(C_0-C_1)$ is equal to $pU(C_0)/(dU/dC)$ where p is the probability of death and $U(C_0)/(dU/dC)$ represents the implicit value of a life. Thus, the compensating variation may be combined with the increase in the probability of death to yield estimates of the value of a

---

*Department of textiles and consumer economics, University of Maryland.

[1] The assumption that the $U(\text{death})=0$ is a convenient simplification. See Hirshleifer, Bergstrom, and Rappaport for a discussion of this assumption.

life. Alternatively, the amount of money an individual is willing to pay for risk reduction may be used to estimate the value of a life.

It should be emphasized that this approach focuses on the rate at which an individual is willing to exchange a marginal change in safety for a marginal change in income. If a large number of individuals are making similar choices, then the estimated value of life to such individuals may be described in the following terms. Suppose 1,000 persons require a compensation of $100 due to a decrease in survival probability of 0.001, then the estimated value of a life is $100 thousand for individuals in the community. Thus, there will be one more fatality once the hazard is introduced and the total required community compensation is $100 thousand. However, this does not mean that any one individual will be willing to sacrifice his life for $100 thousand (see Hirshleifer, Bergstrom, and Rappaport).

II. Methodology

Consideration was given to the reduction in the probability of death and hospitalized injuries due to purchases of smoke detectors.[2] Smoke detectors were selected since they are the most feasible and lowest cost fire protection and control strategy for households. The cost of smoke detectors was then compared to the weighted change in the probabilities of death and injuries. It was assumed that subjective probabilities (which underlie the expected utility approach) might be approximated by relative frequencies. This assumption has been used in other empirical studies, notably by Thaler and Rosen. Its justification in the case of smoke detectors is based on increased public awareness of hazards, particularly life safety hazards, from residential fires due to safety education programs by the National Fire Protection Association (NFPA), National Fire Prevention and Control Administration (NFPCA), U.S. Consumer Product Safety Commission (CPSC) and the promotion of smoke detectors as a method for reducing the consequences of residential fires.

A modification of a method developed by Lester Lave and Warren Weber was used to include the impact of smoke detectors on deaths and hospitalized injuries. The formula for estimating "b," the value of a life, is given as

$$z = bw'x, \; b = (z)(w'x)^{-1} \tag{1}$$

where z = annualized cost of smoke detectors based on capital recovery factor; b = value of a life; w' = weighting vector representing the individual's attitudes towards death or hospitalized injury; and x = probability vector representing changes in the probability of death or hospitalized injury due to purchases of smoke detectors.

It was assumed that smoke detectors have a life expectancy of ten years and that the average household with smoke detectors owned 1.5 smoke detectors (see R. Thompson). Costs of smoke detectors included purchase price and annual replacement cost of batteries. Purchase price was based on catalog prices of Sears, Roebuck and Company. Two discount rates (5 and 10 percent) designed to reflect the subjective opportunity cost of saving or borrowing to the consumer were used to estimate annualized costs.

Three weighting schemes which might be considered by a purchasing household are given in Table 1. The use of a range of weighting schemes was necessitated by the fact that no information is available concerning the relative importance attached by households to the probabilities of death or injury. The first scheme assigns the greatest weight to injury (50 percent) compared to weighting schemes 2 and 3, where values of 10 percent and 0 are assigned. The last scheme means that the individual attaches

---

[2] Minor nonhospitalized injuries were excluded since it was believed that probability reductions for such injuries would have little impact on the decision to purchase smoke detectors. The remaining residential fire loss—property damage—was also excluded since such losses are covered by insurance. Thus, the emphasis was on the role of smoke detectors in averting deaths or serious injuries which is the major theme of most fire safety educational campaigns.

TABLE 1—WEIGHTING SCHEMES

Type of Injury	Scheme 1	Scheme 2	Scheme 3
Fatal	1.0	1.0	1.0
Nonfatal	0.5	0.1	0.0

no value to the reduction in the probability of injury and is only concerned with a reduction in the probability of death. In contrast, death is viewed as only twice as grave as injury in weighting scheme 1. Scheme 3 may be the most appropriate weighting scheme for smoke detectors since educational campaigns have focused on the lifesaving role of smoke detectors. Both weighting schemes 2 and 3 are more realistic than scheme 1 which was introduced primarily to examine the sensitivity of the analysis to variations in weighting schemes. Most empirical studies have used weighting scheme three (see Acton, S. J. Melinek, and Thaler and Rosen).

The change in the probability of death or injury was based on the estimated number of residential fire deaths and injuries, operating rates for smoke detectors and the degree of protection provided by smoke detectors. It was estimated that there were approximately 6,200 residential fire deaths and 6,750 residential fire injuries in 1976. Fire death estimates were based on reports by the NFPA (see L. Derry, p. 50) while fire injury estimates were based on reports by the NFPCA and CPSC, and J. I. Barancik and M. A. Shapiro. The low number of hospitalized injuries relative to deaths is typical and reflects the nature of the hazard posed by residential fire. Occupants either escape or perish.

Data for 1976 reflect the fact that 13 percent of households in 1976 were equipped with smoke detectors. However, not all of these smoke detectors were operational. According to a recent study, only 80 percent of installed smoke detectors may be deemed operational at one point in time (see S. G. Helzer, B. Buchbinder and F. L. Offensend). Household usage or saturation rate (13 percent) multiplied by the operating rate (80 percent) yielded an effective saturation rate of 10 percent for all households. It has also been estimated that smoke detectors will provide 45 percent protection against death and 30 percent protection against injury (personal communication with R. W. Bukowski; National Bureau of Standards). The reason for the relatively low degree of protection is due to a variety of factors including inability to escape, inability to respond to an alarm, and failure to respond correctly.

When the effective saturations rate for smoke detectors was combined with the degree of protection, it was estimated that there would have been 6,492 deaths and 6,759 injuries in 1976 in the absence of smoke detectors. Provision of smoke detectors in all households would have resulted in the avoidance of 2,337 deaths ($0.8 \times 0.45 \times 6492$) and 1,670 injuries ($0.8 \times 0.30 \times 6759$). These numbers were combined with the number of households (74 million) to yield probability reductions of 3.16×10^{-5} and 2.26×10^{-5} for deaths and injuries, respectively.

The above estimates were based on the assumption that residential fires are perceived to be single incident fires, that is, involving one death or one injury. While data from the NFPCA support the dominance of single incident fires, particularly in the case of death, it is possible that multiple person fire incidents may have been envisaged by some purchasers of smoke detectors. In this instance, the resulting b values are overestimated.

Annualized costs of smoke detectors were combined with weighted probability reductions to yield estimates of the value of life to purchasing households from 1974 to 1978. The resulting estimates were combined with sales of smoke detectors to provide estimates of the value of life to the total population at risk.

III. Results

Sales and annualized costs of smoke detectors from 1974–79 are given in Table 2. While some smoke detectors were purchased prior to 1974, it is believed that they account for a negligible proportion of smoke detec-

TABLE 2—ANNUALIZED COSTS OF SMOKE DETECTOR PER HOUSEHOLD ($1976)

Time Period	Sales[a] (Million)	Purchase Price	Annual Operating Costs	Annualized Costs[b]	
				5% Discount Rate	10% Discount Rate
1974	1.8	$51.69	$7.00	$19.15	$21.37
1975	4.0	50.86	7.00	18.99	21.17
1976	7.5	39.65	1.57	9.75	11.65
1977	11.5	25.74	1.57	7.04	8.25
1978	10.0	17.40	1.57	5.42	6.21
1979	10.0	11.85	1.57	4.34	4.86

Sources: National Bureau of Standards (see Bukowski) and Thompson.

[a] Excludes sales mandated by state or local ordinances.

[b] Assuming 1.5 smoke detector per household.

TABLE 3—ANNUAL *b* VALUES FROM PURCHASES OF SMOKE DETECTORS (5 Percent Discount Rate)

Time Period	Weighting Scheme 1	2	3
1974	$446,387	$564,897	$606,013
1975	442,659	560,177	600,949
1976	227,273	287,611	308,544
1977	164,102	207,670	222,785
1978	126,340	159,882	171,519
1979	101,165	128,023	137,342

TABLE 4—ANNUAL *b* VALUES FROM PURCHASES OF SMOKE DETECTOR (10 Percent Discount Rate)

Time Period	Weighting Scheme 1	2	3
1974	498,135	630,383	676,266
1975	493,473	624,484	669,937
1976	271,562	343,658	368,671
1977	192,308	243,363	261,076
1978	144,755	183,186	196,519
1979	113,287	143,363	153,797

tor sales for the time period up to and including 1979. Annualized costs decline considerably from 1975 reflecting technological advances in the production and operating costs of smoke detectors. Sales increased initially but have stabilized in recent years. The value of a life to households purchasing smoke detectors from 1974 to 1979 are given in Tables 3 and 4 for discount rates of 5 and 10 percent, respectively. The *b* values decline over time reflecting a decline in annualized costs over time and a corresponding decline in the value of life by purchasing households. The highest *b* values are observed for weighting scheme 3, which is expected since the only anticipated benefit from the purchase of a smoke detector is the reduction in the probability of death. However, the difference between the three weighting schemes does not appear considerable. In particular, values for weighting schemes 2 and 3 are close.

A comparison of Tables 3 and 4 indicates the impact of the discount rate on the estimated *b* values. Values increase as the discount rate increases. This contrasts with life valuation estimates based on the human capital approach. However, the willingness-to-pay approach is concerned with cost annualization over the product life cycle, while the human capital approach is concerned with foregone earnings over the life of the individual. There is, therefore, no reason to expect the two measures to react similarly to changes in the discount rate.

The data in Tables 3 and 4 may be interpreted two ways: 1) the value of a life required to justify the purchase of a smoke detector; 2) the value of life to an individual who purchases smoke detectors. The second interpretation, which is of concern to this study, requires the assumption that subjective probabilities may be approximated by relative frequencies. Thus, the usage of

TABLE 5—AVERAGE *b* VALUES FROM PURCHASES OF SMOKE DETECTORS: PURCHASING HOUSEHOLDS

Time Period	Discount Rate (%)	Weighting Scheme 1	Weighting Scheme 2	Weighting Scheme 3
1974–79	5	189,049	239,239	256,652
	10	217,273	274,956	294,968

smoke detectors by households will reduce the total number of residential fire deaths. A change in the number of fire deaths per household is obtained by dividing the estimated reduction in fire deaths by the total number of households. If the purchaser of smoke detectors is aware of this change in the number of expected deaths per household, then the price paid for smoke detectors may be used to estimate the value of a life. In view of the greater publicity given to residential fire deaths as opposed to residential fire injuries, the approximation of subjective probabilities by relative frequencies is more likely to be justified for weighting scheme 3.

Data on smoke detector sales in 1974–79 were used to estimate the value of a life to purchasing households in this time period. The value of a life for each time period was multiplied by the relative importance of sales in each time period to yield a weighted estimate of the value of life to purchasing households. Such households comprised 40 percent of all households in the 1974–79 time period. The results are given in Table 5. Values range from $189,049 for weighting scheme 1 and a 5 percent discount rate to $294,968 for weighting scheme 3 and a 10 percent discount rate. However, these results provide an overestimate of the value of a life to the total population at risk since 60 percent of the households are not represented. Jones-Lee comments "the value...of an anticipated saving of one (anonymous) life during the forthcoming periods is given by the average over the relevant population of the marginal value of a decrease of risk" (p. 130).

Assuming that the nonpurchasing population is informed of the risks from household fires, data for 1979 might be used to infer an upper bound to life values for this sector of the population. Thus, in the case of weighting scheme 3, life values are less than $137,342 and $153,797 for discount rates of 5 and 10 percent, respectively.

IV. Discussion

The results obtained in this study fall within the range obtained by Acton, and by Thaler and Rosen. Acton used the questionnaire method and queried individuals concerning the willingness to pay for a reduction in the probability of a fatal heart attack. He obtained values ranging from $41,000 to $63,000 ($1976). Thaler and Rosen used risk premiums paid to workers in hazardous occupations and used regression analysis to hold the impact of other variables constant. Their estimate was $300,000 ($1976), which is considerably higher than the results obtained by Acton.

While the results of this study appear reasonable, certain methodological problems underlying the analysis must be considered. First, the use of relative frequency to approximate subjective probability is crucial to the analysis. While the procedure is necessitated by the fact that subjective probabilities are nonobservable, it may or may not be justified. Jones-Lee argues that "there seems to be no reason to suppose, that, over a large sample of individuals, there will be a systematic divergence between subjective probabilities, and relative frequencies when information concerning the latter is available to the individual decision-maker" (p. 153). The degree of information possessed by the individual then becomes important. The questionnaire method avoids this problem by providing frequency data to the respondents, though a new problem—the validity of answers to hypothetical questions—is created (see Gary Fromm, pp. 166–76).

The second area of concern is the calibration of probabilities by the individuals. Thus, the individual is required to react to small changes in small probabilities and to distinguish between probability reductions of 2×10^{-6} and 4×10^{-6}. This is a risk perception/risk response problem which has not been sufficiently addressed. Yet, it is fundamental to the entire willingness-to-pay approach. Finally, the weighting scheme is also subjective. The relative unimportance of weighting schemes used in this study reflected the fact that the probability reductions for death and injury were of similar orders of magnitude. However, for many consumer products, injuries will far outweigh deaths and the particular weighting scheme selected may then become important.

In view of the problems associated with estimating the value of life, it might be queried whether an alternative approach to life valuation might be used. Thus, the methodology outlined in this paper may also be used by public policymakers to determine the value of a life implicit in their decisions. For example, a program which requires installation of smoke detectors in all new construction for an annualized cost of $4.34 has implicit life values ranging from $101,000 to $137,000 (5 percent discount rate). These values could then be compared to values implicit in other proposed health and safety programs. While this approach does not resolve the controversial issue of life valuation, it does provide decision makers with a basis for program comparisons. Selection of projects with lower implicit life values would, in turn, increase the number of lives that might be saved for a given expenditure of resources and hence, the effectiveness of the entire health and safety effort.

REFERENCES

J. P. Acton, "Evaluating Public Programs to Save Lives: The Case of Heart Attacks," R-950-RC, Rand Corp., Jan. 1973.

J. I. Barancik and M. A. Shapiro, "Pittsburgh Burn Study," unpublished paper, Environ. Health Programs, dept. publ. health practice, Grad. Sch. Publ. Health, Univ. Pittsburgh 1972.

B. C. Conley, "The Value of Human Life in the Demand for Safety," *Amer. Econ. Rev.*, Mar. 1976, *66*, 45–55.

L. Derry, "A Study of United States Fire Experience, 1976," *Fire J.*, Nov. 1977, *71*, 50–53.

G. Fromm, "Comments [on Schelling]," in Samuel B. Chase, Jr., ed., *Problems in Public Expenditure Analysis*, Washington 1969.

S. G. Helzer, B. Buchbinder, and F. L. Offensend, "Decision Analysis of Strategies for Reducing Upholstered Furniture Fire Losses," tech. note no. 1101, National Bureau of Standards, Washington 1979.

J. Hirshleifer, T. Bergstrom, and E. Rappaport, "Applying Cost-Benefit Concepts to Projects which Alter Human Mortality," unpublished paper, Univ. California-Los Angeles, Nov. 1974.

M. W. Jones-Lee, *The Value of Life: An Economic Analysis*, Chicago 1976.

L. B. Lave and W. E. Weber, "A Benefit-Cost Analysis of Auto Safety Features," *Appl. Econ.*, No. 4, 1970, *2*, 265–75.

J. Linnerooth, "The Evaluation of Life-Saving: A Survey," Int. Instit. Appl. Systems Anal., Laxenburg, July 1975.

S. J. Melinek, "A Method for Evaluating Human Life for Economic Purposes," *Accident Anal. Prevention*, Oct. 1974, *6*, 103–14.

E. J. Mishan, *Cost-Benefit Analysis*, New York 1976.

T. C. Schelling, "The Life You Save May Be Your Own," in Samuel B. Chase, Jr., ed., *Problems in Public Expenditure Analysis*, Washington 1969.

R. Thaler and S. Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in Nestor Terleckyj, ed., *Household Production and Consumption*, New York 1975.

R. Thompson, "Cost-Benefit Analysis of Two Alternative Strategies for Reducing Residential Fire Losses," unpublished masters thesis, Univ. Maryland 1978.

D. Usher, "An Inputation to the Measure of Economic Growth for Changes in Life Expectancy," in Milton Moss, ed., *Measurement of Economic and Social Performance*, New York 1973.

R. J. Zeckhauser, "Procedures for Valuing Lives," *Publ. Pol.*, Fall 1975, *23*, 419–64.

National Fire Prevention and Control Administration, *Highlights of Fire in the United States*, Washington, June 1978.

National Fire Protection Association, *Fire Protection Handbook*, 14th ed., Washington 1976.

U.S. Consumer Product Safety Commission, "National Household Fire Survey," Washington 1974.

Factor-Price Uncertainty with Variable Proportions: Note

By STYLIANOS PERRAKIS*

Marion Stewart recently examined the theory of the neoclassical firm under factor-price uncertainty in the same spirit as the well-known Leland-Sandmo theory of the firm under demand uncertainty. The main finding of his study is that, in most cases, risk-averse managers when faced with input-price uncertainty would substitute fixed (riskless) factors of production for the risky factors. The principal methodology used consisted of a direct comparison of the input choices of the firm's manager under both risk neutrality and risk aversion.

Stewart's results are important and have many obvious empirical implications, but they contain two serious deficiencies that may lead to errors and confusions. First, the formulation and solution of the multiple input case in the Appendix, as well as the uncertain price case in Section IV are flawed, insofar as the decision rules followed by the firm are suboptimal. Second, the first paragraph of Stewart's study implies that a risk-indifferent manager operating in an uncertain environment is not going to be affected by the uncertainty with respect to his output and factor-proportions decisions. This implication is misleading. It is shown in this paper that input price uncertainty combined with risk indifference has a very definite effect upon the input choices of the firm as compared to the certainty case. This effect of uncertainty combined with risk neutrality must be distinguished from that of risk aversion proper because it has direct implications for empirical work.

In empirical work at the firm or industry level, input prices are estimated by dividing the observed *ex post* cost of each input by the observed input use during the reference period. The result is an observed "average" price that, under certain assumptions, corresponds to an unbiased estimate of the *expected ex ante* input price. If (as it often happens) this observed price is used in the empirical work as if it were known *ex ante* with certainty, then in general a bias will be introduced *even when the manager is risk indifferent*. It is shown in this paper that in cases of practical interest this bias may or may not have the same sign as the effect of risk aversion, which means that the total effect of uncertainty upon input choices is very often unpredictable when compared to the predictions of the theory of the firm under certainty if the input and output prices are assumed equal to their mathematical expectations. This weakens somewhat the impact of Stewart's results.

The other point made in this note is that in cases of more than one "risky" input, the method followed by Stewart (in his Appendix A) is incorrect, insofar as it leads to suboptimal decisions. Under variable proportions production the firm does not need to choose simultaneously the levels of all inputs. Hence, the risky inputs can be purchased *after* uncertainty has already been resolved, given the level of the riskless input. The selection of the latter must therefore incorporate the fact that risky inputs have been selected *ex post* under conditions of certainty. Similar remarks are also true when there is combined uncertainty in input and output prices.

I shall consider a hypothetical firm that uses three[1] inputs, x_1, x_2, and x_3, to produce the final product q, and let r_1, r_2, r_3, and p denote their respective unit prices. Input x_3 (plant and equipment) must be chosen prior to the other two, so that at the time x_3 is

---

*Associate professor, Faculty of Administration and department of economics, University of Ottawa.

[1] If there is only one variable input, as in Stewart, then the effect of uncertainty (as distinct from that of risk aversion) disappears. See below.

chosen the prices r_1 and r_2 that will be paid are not known with certainty. The manager possesses a subjective joint probability distribution for r_1 and r_2, and let $\bar{r}_1, \bar{r}_2$ denote the corresponding mathematical expectations. I will not impose any structure upon the joint distribution of r_1 and r_2, except to assume that one is not proportional to the other, that is, that the price ratio r_1/r_2 is not constant. For instance, r_1 may be raw materials and r_2 operating labor and it is not unreasonable to assume that each one of these inputs will have its own sources of randomness.

Initially p, q, and r_3 are considered nonrandom, and it is assumed that the firm maximizes the expectation of its profits, $E(\pi)$, where

$$\pi \equiv pq - \sum_{i=1}^{3} r_i x_i$$

It should be noted, however, that inputs x_1 and x_2 are chosen under conditions of certainty, since substitution was assumed and the purchase price is known at the time of the purchase. Hence, x_1 and x_2 are chosen optimally *given* the fixed level of x_3.

I. Input Choices under Factor-Price Uncertainty

Given p, q, and the chosen level of the riskless input x_3, the manager chooses x_1 and x_2 for the realized values of r_1 and r_2. Define the firm's variable cost function as

$$(1)\quad C(q, r_1, r_2, x_3) \equiv \operatorname*{Min}_{x_1, x_2} \{r_1 x_1 + r_2 x_2 \mid q \leqslant f(x_1, x_2, x_3)\}$$

corresponding to the production function f. Modern duality theory has established the uniqueness of the correspondence between f and C, and has shown[2] that C is nondecreasing in q, nonincreasing in x_3, linear homogeneous increasing and concave in r_1 and r_2, and convex in q and x_3 whenever the production function exhibits nonincreasing returns to scale. At the time x_3 is chosen, the functional form of C is known but its value is random, depending on r_1 and r_2.

Under risk neutrality, the level of the *ex ante* input x_3 is going to be chosen by maximizing the expected profit $E(\pi)$. Contrary to Stewart's assertion, this choice is *not* going to be the same as the one under certainty for $r_1 = \bar{r}_1$ an $r_2 = \bar{r}_2$. The only reason Stewart's result was obtained was because there was only one variable input, say, x_1. In such a case, the variable cost function is linear in r_1: if $g(x_3, q)$ in Stewart's notation denotes the input x_1 necessary to produce output q given x_3, we have that

$$(2)\quad E[\pi] = E[pq - r_1 g(x_3, q) - r_3 x_3] = pq - \bar{r}_1 g(x_3, q) - r_3 x_3$$

whose maximization with respect to x_3 would obviously yield the deterministic result. By contrast, when there are two variable inputs, the expected profit becomes

$$(2')\quad E[\pi] = E[pq - C(q, r_1, r_2, x_3) - r_3 x_3] = pq - E[C(q, r_1, r_2, x_3)] - r_3 x_3$$

The function inside the expectation is no longer necessarily linear in the random prices r_1 and r_2. In fact, we know that the concavity of C in (r_1, r_2) implies by Jensen's inequality that $C(q, \bar{r}_1, \bar{r}_2, x_3) \geqslant E[C(q, r_1, r_2, x_3)]$. The equality sign holds under special circumstances, for instance when C is linear in r_1 and r_2 (as in the fixed-coefficients production function) or when r_1 and r_2 are strictly proportional to each other. Barring these special cases, however, C is going to be strictly concave in (r_1, r_2), and $C(q, \bar{r}_1, \bar{r}_2, x_3)$ is going to be strictly greater than $E(C)$.

The choice of x_3 will satisfy the following first-order conditions for the maximization of (2′).

$$(3)\quad r_3 = -\frac{\partial E[C]}{\partial x_3}$$

By contrast, under certainty for $r_i = \bar{r}_i$, $i = 1, 2$ this condition becomes $r_3 = -\partial C(q, \bar{r}_1, \bar{r}_2, x_3)/\partial x_3$.

---

[2] See, for instance, W. M. Gorman and Ronald Shephard.

Unfortunately, the concavity in r_1 and r_2 of the variable cost function is not preserved under differentiation and it is not possible to infer the impact of uncertainty upon the choice of x_3 without restricting somewhat the range of production functions. Suppose that we assume, however, that the variable cost function is separable under the following form:

$$(4) \quad C(q, r_1, r_2, x_3) = c(q, x_3) W(r_1, r_2)$$

In Appendix A it is shown that this separability of the cost function corresponds to the production function $f(x_1, x_2, x_3)$ having the form $\Phi[x_3, h(x_1, x_2)]$, where Φ has the properties of a two-input production function and h is a linear homogeneous production function in (x_1, x_2). This separability and homotheticity restriction is satisfied by many functions commonly used in empirical work, such as the Cobb-Douglas and Leontief functions, as well as the well-known three-input generalizations of the conventional constant elasticity of substitution (*CES*) function, by Sato and Uzawa, respectively,

$$q = \left\{ \left[\delta x_1^\rho + (1-\delta) x_2^\rho \right]^{\mu/\rho} + \gamma x_3 \right\}^{\nu/\mu}$$

and $q = \left[\alpha x_1^\rho + \beta x_2^\rho + \gamma x_3^\rho \right]^{\nu/\mu}$

If (4) holds for the variable cost function, then $\partial C / \partial x_3$ is concave in (r_1, r_2) and for all levels of q and x_3

$$-\frac{\partial E(C)}{\partial x_3} \geqslant -\frac{\partial C(q, \bar{r}_1, \bar{r}_2, x_3)}{\partial x_3}$$

Hence, the level of x_3 satisfying (3) is greater than or equal to the level that would exist if the manager operated on his long-run cost function evaluated at the expected input prices. This increased relative use of the fixed input is in addition to and independent from the one found by Stewart, which was solely due to risk aversion.

II. Price and Output Uncertainty

In assuming that price and output are known with certainty while input prices r_1 and r_2 are still unknown at the time x_3 is chosen, we implicitly assume either some form of price regulation or contracts with "long" delivery lags that stipulate both price and output before production has taken place. An alternative assumption is that output price is also random at the time x_3 is chosen. If this happens, however, it makes little sense in most cases to assume that output is determined[3] simultaneously with x_3, given variable proportions production and *ex post* choice of x_1 and x_2. If we adopt the most plausible assumption, namely that x_1, x_2, *and* q are chosen after the values of r_1, r_2, and p have been revealed, then for competitive firms we reach different conclusions from those of Stewart: while uncertainty and risk neutrality cause in most cases a choice of a larger riskless factor than would have been the case if input and output prices had been set equal with certainty to their expectations, risk aversion *decreases* in these same cases the amount of the riskless factor below the amount used by risk-neutral firms.

If for a competitive firm q, x_1, and x_2 are chosen under certainty *given* x_3, then we can define the gross profit function as

$$(5) \quad \Pi(p, r_1, r_2, x_3)$$

$$\equiv \underset{q, x_1, x_2}{\text{Max}} \; \{ pq - r_1 x_1 - r_2 x_2 \mid q \leqslant f(x_1, x_2, x_3) \}$$

which, by duality theory, is increasing in p, decreasing in r_1 and r_2, linear homogeneous and convex in (p, r_1, r_2), and nondecreasing and concave in x_3 whenever the returns to scale are nonincreasing.[4] An application of Jensen's inequality shows clearly that $E[\Pi(p, r_1, r_2, x_3)] \geqslant \Pi(\bar{p}, \bar{r}_1, \bar{r}_2, x_3)$, where $\bar{p}$ is the expectation of p and the expectation in the left-hand side is with respect to p, r_1, and r_2. Under risk neutrality the choice of x_3 takes place by maximizing expected *total* profit $E[\Pi(p, r_1, r_2, x_3)] - r_3 x_3$. The first-order condition is

---

[3] Such an assumption is clearly invalid under perfect competition. It is also inconsistent with the conventional types of regulated monopoly, but it may be justified under some forms of collusive oligopoly.

[4] The term "gross profit" is due to Gorman, whose work also contains the proof of the properties of the profit function.

$$\frac{\partial E[\Pi]}{\partial x_3} - r_3 = 0 \tag{6}$$

As with the cost function, the convexity of the gross profit function is not always preserved under differentiation with respect to x_3. A separability condition that would preserve convexity is for the gross profit function to have the form $\Pi(p, r_1, r_2, r_3) = s(p, r_1, r_2)g(x_3)$. In Appendix B, it is shown that this condition is equivalent to $f(x_1, x_2, x_3)$ having the form $F[g(x_3), x_1, x_2]$, where g is any increasing and concave function and F is a linear homogeneous production function in (g, x_1, x_2). This condition is trivially satisfied by all production functions with constant returns to scale, but also by many other functions of practical interest such as the generalized Cobb-Douglas and some forms of the translog function. If $\partial E(\Pi)/\partial x_3$ is convex in (p, r_1, r_2) then

$$\frac{\partial E(\Pi)}{\partial x_3} \geqslant \frac{\partial \Pi(\bar{p}, \bar{r}_1, \bar{r}_2, x_3)}{\partial x_3}$$

and the concavity of the gross profit function with respect to x_3 implies that the level of x_3 satisfying (6) is greater than or equal to the level that would exist if the manager operated as if the input and output prices were with certainty equal to their expected values.

The interesting result in this case is the fact that, even when uncertainty combined with risk neutrality raises the level of the fixed factor, risk aversion acts in the opposite way and reduces the fixed factor below the risk-neutral level. To see this, note that the maximization of expected utility of profit yields as a first-order condition

$$E\left[U'\left(\frac{\partial \Pi}{\partial x_3} - r_3\right)\right] = 0 \tag{7}$$

where $U' \equiv U'(\Pi - r_3 x_3)$ is a decreasing function of total profit because of the assumed risk aversion. Following Stewart, I set $A \equiv U'(\Pi - r_3 x_3)$, $B \equiv (\partial \Pi/\partial x_3) - r_3$ and observe that the sign of $Cov(AB)$ is *negative:*[5] for $\Pi = \pi(p, r_1, r_2)g(x_3)$, B is increasing in the random factor Π, while A is decreasing in Π. Hence, since $E(AB) = Cov(AB) + E(A)E(B)$, for the value of x_3 that satisfies (7) the sign of $E(B)$ is positive (since $E(A) > 0$ always), and the optimal level of the riskless input under risk aversion is smaller than in the risk-neutral case. As a result, the net effect of uncertainty plus risk aversion when compared to certainty is indeterminate.

For an imperfectly competitive market structure with an exogenously determined price and a random quantity revealed *ex post*, the structure of the decision-making problem is similar to that of the previous section. The choice of x_1 and x_2 is made under conditions of certainty, and the necessary condition for the choice of x_3 is again (3), with the difference that the expectation refers now to r_1, r_2 *and q*. Stewart's conclusions about the indeterminacy of the effect of risk aversion upon the choice of x_3 are also valid with respect to the comparison between the certainty, and the uncertainty plus risk-indifference cases.

III. Conclusions

Uncertainty in the prices of some inputs at the time the level of other inputs is chosen distorts the factor-proportions choice of the firm when more than one risky input is involved no matter what the manager's attitude towards risk may be. Under risk neutrality and with nonrandom output price and quantity, the distortion results, in most cases of practical interest, in a heavier relative use of the riskless input as compared to the certainty case with input prices set equal to their expectations. This effect is clearly distinct from the one found by Stewart, which was due solely to risk aversion.

When output price is also random and under competitive conditions the most plausible assumption is that output, as well as risky inputs, are chosen under conditions of certainty. By the use of the gross profit function, it is shown that Stewart's results are no longer valid in this case. Even when risk indifference may cause an increased relative use of the fixed input, risk aversion acts in the opposite direction and the total effect of uncertainty is indeterminate.

---

[5] The covariance between an increasing and a decreasing function of a random variable is always negative. See Theorem 236 of G. H. Hardy, J. E. Littlewood, and G. Polya.

APPENDIX—A

Suppose that the production function has the form $f(x_1, x_2, x_3)=\Phi[x_3, h(x_1, x_2)]$, with Φ having the properties of a two-input production function and h being a linear homogeneous production function in (x_1, x_2). Define the *inverse* of Φ as follows:

$$\text{(A1)}\quad c(q, x_3)=\text{Min}\{h \mid \Phi(x_3, h) \geqslant q\}$$

Following Shephard, also define the level set

$$\text{(A2)}\quad L[c(q, x_3)]=\{(x_1, x_2) \mid h(x_1, x_2) \geqslant c(q, x_3)\}$$

By definition, the variable cost function is given by (1) and the set $\{(x_1, x_2) \mid q \leqslant f(x_1, x_2, x_3)\}$ is the level set. Hence,

$$\text{(A3)}\quad C(q, r_1, r_2, x_3) =\text{Min}\{r_1 x_1+r_2 x_2 \mid (x_1, x_2)\, \varepsilon L[c(q, x_3)]\}$$

But the level set of (A2) corresponds to that of a linear homogeneous production function with $c(q, x_3)$ in place of output. It was shown in Shephard that such production functions correspond to cost functions that are linear in output. Hence, $C(q, r_1, r_2, x_3)=c(q, x_3) W(r_1, r_2)$ where

$$\text{(A4)}\quad q W(r_1, r_2)=\underset{x_1, x_2}{\text{Min}}\{r_1 x_1+r_2 x_2 \mid h(x_1, x_2) \geqslant q\}$$

Similarly, if $C(q, r_1, r_2, x_3)$ has the form (3), with c and W satisfying the appropriate properties of cost functions, then there exists a linear homogeneous production function such that (A4) and (A3) hold. Since $C(q, r_1, r_2, x_3)$ is also defined by (1), the sequence of the proof is reversible and results in the fact that $f(x_1, x_2, x_3)= \Phi[x_3, h(x_1, x_2)]$.

B

If $f(x_1, x_2, x_3)=F[g(x_3), x_1, x_2]$, g is a nondecreasing function and F is a linear homogeneous production function in (g, x_1, x_2) then I define the gross profit function.

$$\begin{aligned}\text{(A5)}\quad \Pi(p, r_1, r_2, x_3) &= \underset{q, x_1, x_2}{\text{Max}}\{pq-(r_1 x_1+r_2 x_2) \mid q \leqslant F[g, x_1, x_2]\}\\ &= \underset{q, x_1, x_2}{\text{Max}}\left\{pq-(r_1 x_1+r_2 x_2) \mid \frac{q}{g(x_3)} \leqslant F\left[1, \frac{x_1}{g(x_3)}, \frac{x_2}{g(x_3)}\right]\right\}\\ &= g(x_3) \underset{Q, y_1, y_2}{\text{Max}}\{pQ-(r_1 y_1+r_2 y_2) \mid Q \leqslant F(1, y_1, y_2)\}\end{aligned}$$

where $Q \equiv q / g(x_3)$, $y_i \equiv x_i / g(x_3)$, $i=1,2$. The last expression within braces in (A5) corresponds to a profit function $s(p, r_1, r_2)$ and, therefore, $\Pi(p, r_1, r_2, x_3)=g(x_3) s(p, r_1, r_2)$. The sequence is reversible exactly and the separability of the profit function corresponds to the indicated form of the production function if we note that, as shown by Gorman, the gross profit function corresponding to a linear homogeneous production function is linear in the fixed input when the latter is a scalar.

C

x_1 = *Ex post* input
x_2 = *Ex post* input
x_3 = *Ex ante* input
r_1 = Price of x_1
r_2 = Price of x_2
r_3 = Price of x_3
$\bar{r}_i$ = Expectation of r_i
$\bar{r}_2$ = Expectation of r_2
q = Output quantity
p = Output price
$\bar{p}$ = Expectation of p
π = Profit of the firm

$C(\,)$ = Variable cost function
$f(\,)$ = Production function
$W(\,)$, $c(\,)$ = Component of the variable cost function
$\Phi(\,)$ = Special form of the production function
$h(\,)$ = Component of the production function
$\delta, \rho, \mu, \nu, \alpha, \beta, \gamma$ = Parameters of the CES function
$\Pi(\,)$ = Gross profit function
$s(\,)$, $g(\,)$ = Components of the gross profit function
$F(\,)$ = Special form of the production function
$U(\,)$ = Utility function
A = Marginal utility
B = Marginal profit minus price of x_3
$L[\,]$ = Level set
Q = Transformed output
y_1 = Transformed x_1
y_2 = Transformed x_2

REFERENCES

W. M. Gorman, "Measuring the Quantities of Fixed Factors," in J. N. Wolfe, ed., *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, Chicago 1968, 141–72.

G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge 1959.

D. M. Holthausen, "Input Choices and Uncertain Demand," *Amer. Econ. Rev.*, Mar. 1976, *66*, 94–103.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

A. Sandmo, "On the Theory of the Competitive Firm under Price Uncertainty," *Amer. Econ. Rev.*, Mar. 1971, *61*, 65–73.

Ronald W. Shepard, *Cost and Production Functions*, Princeton 1953.

M. B. Stewart, "Factor Price Uncertainty with Variable Proportions," *Amer. Econ. Rev.*, June 1978, *68*, 468–73.

A Model of the Parallel Team Strategy in Product Development

By FRED D. ARDITTI AND HAIM LEVY*

In the late 1950's, the idea of employing several teams to attack the development of weaponry systems—now known as the "parallel path strategy" of development—evolved at the Rand Corporation. This emphasis on exploring several technological paths in the development of a product arose in response to the recognition that the expensive part of research and development (*R&D*) is not the exploratory work, but the latter part of the development process (see Benjamin Klein and Richard Nelson, 1959). This proved to be the strategy adopted in the early development stages of the atomic bomb, where five independent teams at various universities pursued the problem of producing large quantities of material capable of producing a self-sustaining reaction. Similarly, the technique was widely applied to the development of aircraft engines during World War II, when it was found that the best way to produce a good engine was to bring several different prototypes to a practical demonstration as quickly as possible, choose that one which exhibited the most favorable characteristics, and work intensively on the selected model to turn it into an effective military device (see Robert Schlaifer and S. D. Heron). In later periods, the Soviet Union employed this management strategy in developing their MIG fighters, while the United States brought this competition in team research to the development of the Thor and Jupiter missiles. The approach has also proved beneficial in cases where a central authority does not initiate the formation of the several teams. The initial stages in the development of color movies and television were undertaken by several private firms, each following its own plan.

While the idea's birth appears to have many fathers (see Nelson, 1959), only he (1961) can be credited with providing a framework in which the costs and benefits of the parallel path strategy could be analytically studied. The problem considered by Nelson is that of achieving the development of the project at minimum cost. His model assumes that n teams are selected to work on the project up to some review point. When the review date arrives, the performances of the teams are compared, and the team that looks best in terms of successfully completing the project at minimum cost is retained while the remaining $n-1$ teams are dropped. The basic assumption of Nelson's characterization of the parallel path model is that the cost of using several teams during this initial stage is small relative to the benefits that accrue from the information gathered; however, as development continues into later stages, the cost of employing multiple teams increases disproportionately to the value of information obtained.

The Nelson model does indeed explain many of the instances in which we observe a firm or government agency employing more than one team on a development project. However, the model does not explain those cases where more than one team works until development is completed or failure is conceded—failure occurring whenever the allocated development funds run out. Nor does it explain why several teams may be employed even if the per team cost during the initial phase is not low relative to the value of information gathered. Stimulated by Nelson's work, we suggest in this paper another model that explains these situations, and we analyze the factors important in determining the optimal number of parallel teams.

---

*University of Florida-Gainesville, and Hebrew University of Jerusalem and the University of Florida-Gainesville, respectively. Levy acknowledges financial support from the University of Florida's Public Policy Institute. We wish to thank an anonymous referee and George Borts for helpful comments and suggestions.

We assume that the size of each team is optimal, and the only question is how many teams should be employed in *R&D*. Obviously, when we determine the number of teams to be employed, we determine the total optimal investment in *R&D*. This issue is crucial in considering the government budget allocated to *R&D* as well as the amount of money invested in *R&D* by the private sector. Just to indicate the magnitude of the investment in *R&D*, note that in 1977, General Motors invested $1,451 million in *R&D*; and the corresponding investment figures for other leading firms are, respectively: Ford Motor Company, $1,170 million; IBM, $1,142 million; and AT&T, $718 million.

Finally, we would like to mention that we suggest in this paper a relatively simple model that can be extended in various directions; for example, flow of information between teams, agreement between competitors working on the discovery of two or more products which either substitute or complement each other, etc. However, we believe that the simple model contains the main message to which we would like to address ourselves.

I

A. *The Model*

The firm is searching for a new product. If the search is successful, the discovered product increases the firm's net worth by its net present value R. In calculating R, all costs to be incurred in the future production of the product being developed are accounted for, except the costs of development, A dollars per team, that are considered separately. Clearly $R>0$, since otherwise no attempt will be made to develop the product.

Given that any team's chances of success are independent of the number of teams used and that the probability of success is $1-q$ while that of failure is q, how many teams should the firm employ during the development stage? If only one team is employed, the expected dollar return $\Pi(1)$ is

$$\Pi(1)=(1-q)R-A \tag{1}$$

where the development cost of A dollars is incurred with certainty. With two teams, the expected return $\Pi(2)$ is the probability that at least one team will succeed, or one minus the probability of both teams failing q^2 times R minus the cost of using two teams. Formally,

$$\Pi(2)=(1-q^2)R-2A \tag{2}$$

Suppose that employment of the first team is profitable, should we add a second team? Only if the addition results in an increase in profit. That is, if

$$\begin{aligned}\Delta\Pi(1)&=\Pi(2)-\Pi(1)\\ &=q(1-q)R-A>0\end{aligned} \tag{3}$$

or

$$q(1-q)>A/R \tag{3'}$$

Obviously, the lower the cost-to-present-value ratio A/R, the more likely it is that a second team will be added. The effect of q on the decision to add another team is rather more complicated, for, given A/R, a decrease in q does not necessarily work in favor of increasing the number of teams. The reason is that a decrease in q gives rise to two counterforces. On the one hand, one can always increase the chances of success by adding a team. The higher q is, the more attractive this strategy is bearing in mind that the increase in odds must outweigh the added certain cost of A dollars. On the other hand, the lower q is, the greater the likelihood that a single team will succeed, dictating against the addition of a second team—since the event "two or more teams discover the new product" yields the same return, R, as the event "only one team discovers the new product." For $q>1/2$, the former force dominates in favor of increasing the number of teams, while for $q<1/2$, the latter force prevails.[1]

---

[1] Note that the marginal profit induced by adding a second team is given by $\Delta(\Pi_1)=q(1-q)R-A$. Thus, $\partial\Delta(\Pi_1)/\partial q=R(1-2q)$, which is always a decreasing function of q. However, for all $q<1/2$, $\partial(\Delta\Pi_1)/\partial q$ is positive and for all $q>1/2$, $\partial(\Delta\Pi_1)/\partial q$ is negative, which explains the above two counterforces. In technical terms these two counterforces stem from the fact that $q(1-q)$ is a parabola, reaching its maximum at the value $q=1/2$.

More generally, if there are $n-1$ ongoing teams, the question of adding an nth team is decided by

$$\Delta\Pi(n-1)=\Pi(n)-\Pi(n-1)>0 \tag{4}$$

where

$$\Pi(n)=(1-q^{n})R-nA \tag{5}$$

Substitution of (5) into (4) yields

$$\Delta\Pi(n-1)=q^{n-1}(1-q)R-A>0 \tag{6}$$

or

$$q^{n-1}(1-q)>A/R \tag{6'}$$

The critical point to focus on is $q=(n-1)/n$, for if q is below this value, then a decrease in q works against adding the nth team while if q is above $(n-1)/n$ then a decrease in q tends to favor an increase in teams.[2]

B. *The Optimal Number of Teams*

If n^* is the number of teams that produces maximum expected return, then it is necessary that $\Pi(n^*)$ satisfy

$$\Delta\Pi(n^*-1)=\Pi(n^*)-\Pi(n^*-1)>0 \tag{7a}$$

$$\Delta\Pi(n^*)=\Pi(n^*+1)-\Pi(n^*)<0 \tag{7b}$$

Employing the relationship given by equation (6) for the cases n^* and n^*-1, the following must hold, if indeed n^* provides the maximum expected profit,

$$Rq^{n^*-1}(1-q)-A>0 \tag{8a}$$

$$Rq^{n^*}(1-q)-A<0 \tag{8b}$$

Equations (8a) and (8b) imply that

$$q^{n^*-1}>\frac{A}{R(1-q)}>q^{n^*} \tag{9}$$

Noting that $\ln q<0$, then n^* must satisfy

$$\left(\frac{1}{\ln q}\right)\ln\left[\frac{A}{R(1-q)}\right]<n^*<1+\left(\frac{1}{\ln q}\right)\ln\left[\frac{A}{R(1-q)}\right] \tag{10}$$

Condition (10) is sufficient for a global maximum, since the second differential is everywhere negative. That is

$$\begin{aligned}\Delta^2\Pi(n)&=\Delta\Pi(n)-\Delta\Pi(n-1)\\&=Rq^{n}(1-q)-A-Rq^{n-1}(1-q)+A\\&=Rq^{n-1}(1-q)(q-1)\\&=-Rq^{n-1}(1-q)^2<0 \qquad \textit{for any } n\end{aligned} \tag{11}$$

From condition (10), the importance of q's size in determining the optimal n is evident. We see that as q increases from zero, there is a point above which $\ln[A/(R(1-q))]$ becomes positive,[3] which in turn forces the upper bound on n^* to fall below one.[4] In this event, the firm views the project as unprofitable and no team is assigned to work on it. The above reasoning also tells us that for at least one team to be employed, we should have $\ln[A/(R(1-q))]<0$, namely $A/R(1-q)<1$ or $R(1-q)-A>0$, i.e., the project must have positive net present value.

Finally, recall that the optimal number of teams n^* is the same under the assumption of certain R, and under the assumption that R is a random variable when the firm's goal is to maximize its expected net present value. This identity becomes transparent once we recall that the net present value is $(1-q^{n})R-nA$, and the expected net present value is given by $(1-q^{n})E(R)-nA$ (see equations (5) and (21)).

---

[2] This is true since the function $q^{n-1}(1-q)$ reaches its maximum at the value $q=n-1/n$ (see also fn. 1).

[3] This critical point is given by $q=1-(A/R)$, where $A/R<1$, otherwise the project is not worth considering.

[4] Recall that $\ln q<0$.

C. *The Change in n^* Due to Parameter Changes*

Let us turn to the analysis of the economic factors which determine the optimal number of teams, n^*. This analysis relies on the inequality given in (10). Define

$$(12)\quad B \equiv \left(\frac{1}{ln\, q}\right) ln\left[\frac{A}{R(1-q)}\right]$$

$$= \left(\frac{1}{ln\, q}\right)[ln\, A - ln\, R - ln(1-q)]$$

From (10), the higher B, the higher n^*. Therefore to understand how a parameter change affects n^*, one needs only study the effect of that parameter change on B. As one suspects,

$$(13)\quad \frac{\partial B}{\partial A} = \frac{1}{A(ln\, q)} < 0; \quad \frac{\partial B}{\partial R} = -\frac{1}{R(ln\, q)} > 0$$

An increase in per team cost or a decrease in project profitability tends to reduce the optimal number of teams.

With regard to a change in q

$$(14)\quad \frac{\partial B}{\partial q} = \frac{1}{[ln\, q]^2}\left[\frac{ln\, q}{(1-q)}\right] - \frac{1}{q} ln\left[\frac{A}{R(1-q)}\right]$$

So the sign of $\partial B/\partial q$ is determined by the sign of the bracketed terms. Because we assume the project is expected to be profitable for at least one team, i.e., $R(1-q)-A>0$, then

$$-\frac{1}{q} ln\left[\frac{A}{R(1-q)}\right] > 0$$

but $\frac{ln\, q}{(1-q)} < 0$

Therefore the effect on n^* of an increase in q is ambiguous. The best one can do is state the condition under which $\partial n^*/\partial q > 0$. From (14), the condition is

$$(15)\quad q\, ln\, q - (1-q)\, ln\left[\frac{A}{R(1-q)}\right] > 0$$

or

$$(16)\quad q^q\left[\frac{R(1-q)}{A}\right]^{1-q} > 1$$

II. Risk Analysis

Even if a team is successful in development, the final product's future income is a random variable, consequently the product's present value, symbolized by R in previous sections, can no longer be assumed known but must be considered a random variable. This uncertainty with respect to the payoff from successful development constitutes one type of risk. The other risk associated with product development is that none of the n^* teams will succeed and no new product will be forthcoming. When the above risks are explicitly considered, what constitutes an optimal strategy? We analyze this issue by employing stochastic dominance criteria and the well-known mean-variance rule.

A. *Stochastic Dominance*

In this section we make use of two theorems on ordering preferences under risk. Both theorems have been grouped under the heading of stochastic dominance. We begin by defining new notation, state the two theorems, and then proceed with the analysis.

Without loss of generality we assume that the firm faces a decision of employing either n_1 or n_2 teams, where $n_2 > n_1$. We refer to a decision to employ n_1 teams as strategy 1 and to a decision to employ n_2 teams as strategy 2. The net income from an investment in *R&D* is a random variable which we denote y. Furthermore, denote $F_1(y)$ and $F_2(y)$, respectively, as the cumulative distribution functions of the net present value of the random variable y, induced by employing n_1 or n_2 teams.

THEOREM 1: *First-Degree Stochastic Dominance (FSD). Given two cumulative probability distributions F_1 and F_2, strategy 1 will be preferred to strategy 2 by every expected utility maximizer, independent of the convexity or concavity of the utility function, if and only if $F_1(y) \leq F_2(y)$ for all values y, and for* *at least one y value the strict inequality $F_1(y) < F_2(y)$ holds.*

(For proof of *FSD* see Josef Hadar and William Russell; Giora Hanoch and Levy, 1969; and James Quirk and Rubin Saposnik.)

THEOREM 2: *Second-Degree Stochastic Dominance (SSD). Given two cumulative probability distributions F_1 and F_2, strategy 1 will be preferred to strategy 2 by every expected utility maximizer with a concave utility function, if and only if*

$$\int_{-\infty}^{y}[F_2(t)-F_1(t)]\,dt \geqslant 0$$

for all y, and the strict inequality holds for at least one y value.

(For proof of *SSD* see Peter Fishburn, Hadar and Russell; Hanoch and Levy, 1969; and Michael Rothschild and Joseph Stiglitz.)

Assume that the probability density function of the project's present value, excluding development costs, is given by

$$(17) \qquad g(R)=\begin{cases} 0 & R\leqslant 0 \\ h(R) & R>0 \end{cases}$$

where $\int_{-\infty}^{\infty} g(R)dR=\int_{0}^{\infty} h(R)dR=1$

If a development program using n teams is unsuccessful, that event having probability q^n, then the project's net present value y equals $-nA$; on the other hand, if the program is successful, the probability of that event being $1-q^n$, then the y random variable is $R-nA$. Since the success (or failure) of the development process is independent of the present value random variable R, the net present value y can be described by the following mixed distribution[5]

$$(18) \quad f(y)=\begin{cases} 0 & y<-nA \\ q^n & y=-nA \\ (1-q^n)h(y+nA) & y>-nA \end{cases}$$

The cumulative distribution of y under the general strategy n is then given by

$$(19) \quad F(y)=\begin{cases} 0 & y<-nA \\ q^n & y=-nA \\ q^n+(1-q^n)H(y+nA) & y>-nA \end{cases}$$

where $\quad H(y+nA)=H(R)=\int_0^R h(t)\,dt$

As n increases, $-nA$ decreases, and the entire distribution $F(y)$ shifts to the left. Moreover, q^n decreases (since $q<1$) and $1-q^n$ correspondingly increases.

Consider three alternative strategies n_1, n_2, and n_3, where $n_3>n_2>n_1$. Using (19) we plot the corresponding cumulative distributions in Figure 1. As we see from Figure 1, no strategy can dominate a strategy with a lower n either by *FSD* (since $F_1(-n_2A)<F_2(-n_2A)$ and $F_2(-n_3A)<F_3(-n_3A)$) or by *SSD*, since

$$\int_{-\infty}^{-n_1A}[F_1(t)-F_2(t)]\,dt<0$$

and $\quad \int_{-\infty}^{-n_2A}[F_2(t)-F_3(t)]\,dt<0$

However, a given strategy n may dominate a strategy with a higher n. Figure 1 illustrates a case where strategy n_1 dominates n_3 by *FSD*, since $F_1(y)\leqslant F_3(y)$ for all y and a strict inequality obtains for $y>-n_3A$. Strategy n_1 may also dominate n_2 by *SSD*, but not by *FSD* (since the two cumulative distributions F_1 and F_2 intersect.) The only way to determine whether strategy n_1 dominates n_2 by *SSD* is to calculate the (algebraic) area between the two curves, that is, to apply the *SSD* rule directly.

In order to calculate the relevant area, one has first to find all the intersection points between the two cumulative distributions under consideration. In Figure 1, we demonstrate a case where F_2 and F_1 intersect only twice. In this case, if the negative area which is between $-n_1A$ and y_0 is smaller than the positive area between the

---

[5] Obviously, $\int_{-\infty}^{+\infty} f(y)dy=q^n+(1-q^n)\times\int_0^{\infty} h(R)dR=q^n+(1-q^n)=1$

Recall that the net present value is $y=R-nA$, hence $R=y+nA$ as stated in equation (18).

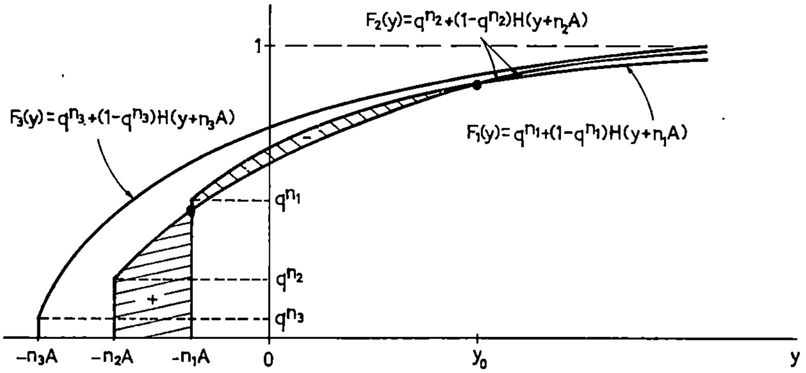


FIGURE 1. THE CUMULATIVE DISTRIBUTIONS OF NET PRESENT VALUE OF STRATEGIES n_1; n_2; $n_3(n_1<n_2<n_3)$

two curves over the range $(-n_2A, -n_1A)$, we can conclude that $F_1(y)$ dominates $F_2(y)$ by *SSD*. Though for some distributions of R (for example, a uniform distribution) it can readily be seen that either there is no intersection at all (for example, $F_3(y)$ and $F_1(y)$ in Figure 1), or if the cumulative distributions cross, then there must be exactly two intersection points (as for $F_2(y)$ and $F_1(y)$ in Figure 1), such assertion is not true in general, and two cumulative distributions may intersect any number of times. Thus, in general, in order to establish dominance one has to analyze the specific distribution of R in the relevant case. Finally, if R is given such that there is only one intersection between the two distributions under consideration, then $F_1(y)$ dominates $F_2(y)$ by *SSD* if and only if the expected net profit by $F_1(y)$ is greater than the expected net profit by $F_2(y)$. The distribution $F_1(y)$ dominates $F_2(y)$ for all concave utility functions since *SSD* is equivalent in this case to the simple rule: the mean of $F_1(y)$ is greater than the mean of $F_2(y)$. In general, given two distributions F_1, F_2 with mean values μ_1, μ_2, respectively, such that for some $y_0 < \infty$, $F_1 \leqslant F_2$ for $y<y_0$ (and $F_1<F_2$ for some $y_1<y_0$) and $F_1 \geqslant F_2$ for $y \geqslant y_0$, then F_1 dominates F_2 (for all concave utility functions, i.e., *SSD*) if and only if $\mu_1 \geqslant \mu_2$. (The proof of this statement is presented in Hanoch and Levy, 1969, p. 371.)

Using some of the properties of *SSD* efficiency, we can determine the efficient set of strategies. A strategy n_i is included in the efficient set if there is no other strategy which dominates it. First, a necessary condition for *FSD* and *SSD* is that the dominated strategy have a lower expected return.[6] Therefore, the strategy that provides the highest expected return, symbolized by n^*, is undominated. Does strategy n^* dominate any other strategy? Recalling the reasoning used in previous paragraphs to prove that strategy $n_2(>n_1)$ cannot dominate n_1, we conclude by the same reasoning that n^* cannot dominate (either by *FSD* or *SSD*) any strategy n where $n<n^*$. Thus, all the strategies $n_i \leqslant n^*$ are *FSD* as well as *SSD* efficient in the sense that there is no strategy which dominates n_i either by *FSD* or by *SSD*.[7] What about strategies $n_j>n^*$? It is obvious that there is no strategy n_j which dominates any of the strategies $n_i \leqslant n^*$, since

---

[6] Let $u(y)=y$, where $u(\cdot)$ denotes utility function, be a member of the class of utility functions for which the *FSD* and the *SSD* theorems hold. If F_1 dominates F_2 by *FSD* or *SSD*, then

$$\int_{-\infty}^{+\infty} y\, dF_1 > \int_{-\infty}^{+\infty} y\, dF_2$$

Thus $\mu_1 > \mu_2$ is a necessary condition for *FSD* and *SSD*.

[7] Since expected profit monotonically increases with n up to n^*, then by the same argument used to prove that n^* cannot be dominated by $n<n^*$ we know that any two strategies n_3 and n_4, $n_3<n_4<n^*$, cannot dominate each other.

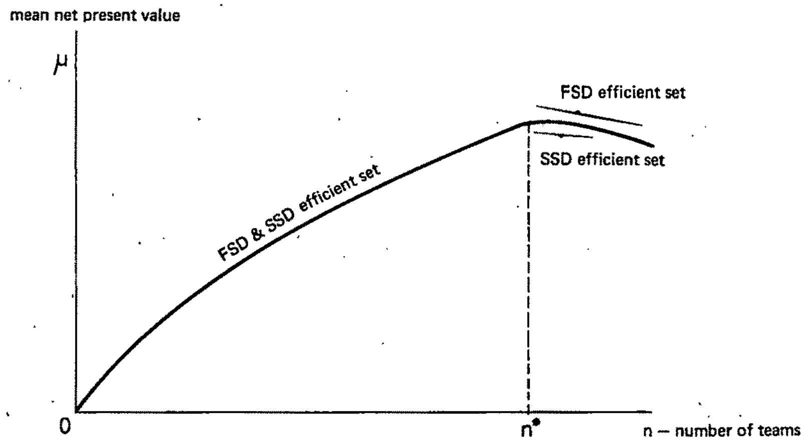


FIGURE 2. EFFICIENCY STRATEGY SET BY FIRST- AND SECOND-DEGREE STOCHASTIC DOMINANCE RULES

the cumulative distribution of n_j starts further to the left than any cumulative distribution of n_i. However, n^* (or any other n_i) may dominate one or all distributions of n_j either by *FSD* or by *SSD*. Such dominance is possible since, on the one hand, n^* is the strategy with the highest expected net present value (this stems from equation (10), since the analysis in the certainty case is equivalent to an analysis involving maximation of the expected net present value), and, on the other hand, the cumulative distribution of n^* starts farther to the right than the cumulative distribution of n_j. In order to see if such dominance exists, one has to apply directly either the *FSD* or the *SSD* rule and compare n^* with all other strategies n_j. However, while all strategies $n_i \leqslant n^*$ are *FSD* (as well as *SSD*) efficient, for $n_j > n^*$ the efficient sets of strategies derived by *FSD* and *SSD* rules do not coincide. Figure 2 illustrates such a possibility when the *SSD* efficient set is smaller than the *FSD* efficient set.

B. *Mean-Variance Analysis*

Almost all of the risk-return efficiency analysis of economic decision making has been done under more restrictive assumptions: 1) the attributes of any risky strategy are completely defined by the strategy's expected return and variance of return; and 2) the decision maker views expected return favorably and variance unfavorably.[8] We now investigate what can be said about the n-efficient set in mean-variance space.

Since the return y from a strategy n is $R-nA$ with probability $(1-q^n)$ and $-nA$ with probability q^n, the expected dollar return y is given by

(20)

$$E(y)=(1-q^n) \times \int_0^\infty (R-nA)h(R)\,dR-q^n\cdot(nA)$$

or

(21)

$$E(y)=(1-q^n)[E(R)-nA]-q^n(nA) = (1-q^n)E(R)-nA$$

---

[8] The mean-variance rule is quite convenient to apply. However, one can safely apply this rule if one either assumes normal probability distributions with risk aversion or quadratic utility functions. (See Harry Markowitz, 1952, 1959; James Tobin, 1958; Nelson, 1961.) It is worth mentioning that in the quadratic utility case the mean-variance criterion is only a sufficient rule. (See Hanoch and Levy, 1970.) In a recent dispute with regard to the theoretical validity of the mean-variance rule, see Levy and Tsiang, and on its empirical approximation to a maximization of a direct expected utility, see Levy and Markowitz.

The variance of $\sigma^2(y)$ is derived as follows:

(22)

$$\begin{aligned}\sigma^2(y) &= \int_0^\infty [R-nA-(1-q^n)E(R)+nA]^2 \\ &\quad \times (1-q^n)h(R)\,dR \\ &\quad +[-nA-(1-q^n)E(R)+nA]^2 q^n \\ &= (1-q^n)\int_0^\infty [(R-E(R)) \\ &\quad +q^nE(R)]^2 h(R)\,dR \\ &\quad +(1-q^n)^2[E(R)]^2 q^n \\ &= (1-q^n)\sigma^2(R)+[E(R)]^2 \\ &\quad \times \left[(1-q^n)q^{2n}+q^n-2q^{2n}+q^{3n}\right] \\ &= (1-q^n)\left[\sigma^2(R)+q^n[E(R)]^2\right]\end{aligned}$$

From the analysis in Section II we know that $E(y)$ reaches a maximum at some n, labelled n^*. What about $\sigma^2(y)$? Partially differentiating $\sigma^2(y)$ with respect to n, we obtain

(23)

$$\begin{aligned}\frac{\partial\sigma^2(y)}{\partial n} &= -q^n(\ln q)\left[\sigma^2(R)+q^nE(R)\right]^2 \\ &\quad +(1-q^n)q^n(\ln q)[E(R)]^2 \\ &= [-q^n \ln q]\sigma^2(R)+q^n(\ln q) \\ &\quad \times [E(R)]^2[1-q^n-q^n]\end{aligned}$$

or

$$\text{(24)}\quad \begin{aligned}\frac{\partial\sigma^2(y)}{\partial n} &= [-q^n \ln q]\sigma^2(R) \\ &\quad +q^n(\ln q)[E(R)]^2[1-2q^n]\end{aligned}$$

Since $\ln q<0$ and $q^n<1$, the first term on the right-hand side of (24) is positive while the second term is ambiguous; consequently the sign of $\partial\sigma^2(y)/\partial n$ in general is ambiguous at any n. However, if $[1-2q^n]$ which appears on the right-hand side of (24) is negative then one can determine that $\partial\sigma^2(y)/\partial n>0$. This implies that $\partial\sigma^2(y)/\partial n>0$ if $1<2q^n$ or $n \ln q>\ln 1/2$. Recalling that $\ln q$ is negative, the condition for a positive sign in (24) is $n<\ln 1/2/\ln q$. Thus, what one has to do is to calculate the positive number $\ln 1/2/\ln q$ which we denote by n^{**}. Compare n^{**} to n^*, which is the strategy with the higher expected return. If $n^{**}>n^*$ one can conclude that in the range $n^*<n<n^{**}$, $E(y)$ decreases and $\sigma^2(y)$ increases with an increase in n, and hence all the strategies of n such that $n^*<n\leqslant n^{**}$ are dominated by strategy n^*.

If, on the other hand, we find that $n^{**}<n^*$, no further general progress can be made toward the derivation of the n-team efficient set in $E(y)-\sigma^2(y)$ space. Nevertheless, given estimates of q, μ_R, and σ_R, one can obtain the efficient (μ_y, σ_y) set by a numerical analysis; the efficient set generated will, of course, vary with the values of the q, μ_R, and σ_R parameters.

III. Concluding Remarks

Working on research and development can be carried out by using many alternative production functions. However, in many cases the firm estimates the number of people and the required resources needed by one investigating team for a predetermined time period. Looking at the costs and the potential revenues that may result from discovery, the firm may decide to go ahead with the project. If this is the case, should the firm add another investigating team? Assuming certainty, or that the firm's objective function is to maximize expected net present value, we solve for the optimal number of teams that should be employed and analyze the impact of changes in the relevant parameters on the optimal number of teams to employ. Obviously, the higher the potential net present value of the new product, apart from development costs, and the lower the investigating cost per team, the higher the optimal number of teams that should be employed. The impact of a change in the probability of discovery of a new product by each team on the optimal solution is somewhat ambiguous: because of counterforces, given two projects identical in all respects except in the probability of success, it is not clear a priori which project should employ the larger number of teams.

Extending the analysis to incorporate uncertainty yields a range of efficient strategies from which the firm should select (subjectively) the optimal strategy. In this analysis we assume either risk aversion and use stochastic dominance criteria, or that the firm works in a mean-variance framework. The first screening of inefficient strategies under the above rules is a function of the estimated present value, the investigating cost per team as well as each team's discovery probability.

Given a subjective function of the potential profit from the new discovery (i.e., $f(R)$), one can use it to derive more explicitly the *SSD* as well as the mean-variance efficient sets of strategies, from which the firm's management should select its optimal strategy and thus determine the optimal investment in research and development.

REFERENCES

Peter C. Fishburn, *Decision and Value Theory*, New York 1964.

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

G. Hanoch and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Rev. Econ. Stud.*, July 1969, *36*, 335–46.

______ **and** ______, "Efficient Portfolio Selection with Quadratic and Cubic Utility" *J. Bus., Univ. Chicago*, Apr. 1970, *43*, 181–89.

B. H. Klein, "A Radical Proposal for R and D," *Fortune*, May 1958, *66*, 112.

H. Levy, "The Rationale of the Mean-Standard Deviation Analysis: Comment," *Amer. Econ. Rev.*, June 1974, *64*, 434–41.

______ **and H. M. Markowitz,** "Approximating Expected Utility by a Function of Mean and Variance," *Amer. Econ. Rev.*, June 1979, *69*, 308–18.

Harry M. Markowitz, "Portfolio Selection," *J. Finance*, Mar. 1952, *6*, 77–91.

______, *Portfolio Selection: Efficient Diversification of Investment*, New York 1959.

T. Marschak, T. Glennon, and R. Summers, "Toward A Normative Theory of Development," in their *Strategy for R&D: Studies in the Microeconomics of Development*, Berlin; Heidelberg; New York 1967.

R. Nelson, "Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts," *Rev. Econ. Statist.*, Nov. 1961, *43*, 351–64.

______, "The Economics of Invention: A Survey of the Literature," *J. Bus., Univ. Chicago*, Apr. 1959, *32*, 101–27.

J. P. Quirk and R. Saposnik, "Admissibility and Measurable Utility Functions," *Rev. Econ. Stud.*, Feb. 1962, *29*, 140–46.

M. Rothschild and J. E. Stiglitz, "Increasing Risk: I. A Definition," *J. Econ. Theory*, Sept. 1970, *2*, 225–43.

Robert Schlaifer and S. D. Heron, *The Development of Aircraft Engines and Fuels*, Cambridge, Mass. 1950.

J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. Econ. Stud.*, Feb. 1958, *26*, 65–86.

______, "The Theory of Portfolio Selection," in F. H. Hahn and F. P. R. Brechling, eds., *Theory of Interest Rates*, New York 1965.

S. C. Tsiang, "The Rational of the Mean Standard Deviation Analysis, Skewness Preference, and the Demand for Money," *Amer. Econ. Rev.*, June 1972, *62*, 354–71.

Löschian Competition under Alternative Demand Conditions

By BRUCE L. BENSON*

Theoretical examination of the pricing behavior of firms competing in a spatial setting indicates that many of the conclusions of classical price theory are altered as the cost of distance becomes significant.[1] An assumption about entrepreneurial behavior is an integral component of these models of spatial competition (see Greenhut, Hwang, and Ohta). The traditional spatial model follows the Löschian assumption in location theory—that entrepreneurs price as if they are monopolists within their market areas.[2] Recently however, this Löschian assumption has been questioned.[3] For example, Greenhut, Hwang, and Ohta noted that unexpected conclusions derived in spatial price theory stem, in part, from this behavioral reaction assumption. Capozza and Van Order further contended that the Löschian assumption does not accurately depict entrepreneurial behavior. They *concluded* that another assumption concerning entrepreneurial behavior "... is closer to the way firms behave in most real world situations" (p. 896).

Capozza and Van Order proposed to investigate the extent to which spatial price theory replicates the results of spaceless price theory with free entry. To this end, they offered five "... intuitively appealing properties [or characteristics] of nonspatial competitive theory..."(p. 897) which they felt a reasonable model of spatial competition should obey:

> 1) As transport costs approach zero, perfect competition should be approached and price should approach marginal cost.
> 2) As fixed costs approach zero, concentrated production is less essential; spatial monopoly power is diminished; and again price should approach marginal cost... .
> 3) As costs (fixed, marginal, or transportation) rise, price should rise.
> 4) As demand density rises, firms should be able to take advantage of economies of scale. Price should fall in the long run.
> 5) As more firms enter the industry, there should be increased competition and price should fall. [p. 897]

Capozza and Van Order analyzed three models of spatial competition with these characteristics as the criteria for acceptance or rejection. They apparently felt that entrepreneurial behavior was *the crucial assumption* in a spatial competition model, since, while developing their models, they proposed: "The final assumption, *upon which our results depend*, concerns the reaction of a firm to a change in a competitor's price" (p. 898, emphasis added). In fact, the only difference between their three models was the entrepreneurial behavior assumption employed in each. Capozza and Van Order claimed that Löschian competition violates all the above characteristics![4] Actually, however, the characteristics can be derived

---

*Assistant professor of economics, Pennsylvania State University. I wish to thank M. L. Greenhut and an anonymous referee for helpful comments and suggestions.

[1] For example, see Melvin Greenhut, Ming Jeng Hwang and Hiroshi Ohta; Dennis Capozza and Kazem Attaran; Capozza and Robert Van Order; and my forthcoming paper.

[2] For example, see Edwin Mills and Michael Lav; Greenhut (1971); John Hartwick; Nicholas Stern; among others.

[3] For example, see Greenhut, Hwang, and Ohta; Capozza and Van Order.

[4] Capozza and Van Order argued that the appropriate assumption is that each entrepreneur believes that all other entrepreneurs will leave their own location and price unchanged—the Hotelling-Smithies assumption. They rejected both Löschian and Greenhut-Ohta behavior. The Greenhut-Ohta assumption is that each entrepreneur assumes the price at the edge of his market area is fixed (see Greenhut, Hwang, and Ohta).

for a Löschian competitor simply by altering another assumption used by Capozza and Van Order (C-V).

The purposes of the following presentation are to demonstrate that: 1) the C-V rejection of Löschian competition is not warranted, at least for the reasons they presented; and 2) the entrepreneurial behavior assumption is not the only crucial assumption in a model of spatial competition. The Löschian assumption produces results which violate the C-V characteristics, only when used with the assumption they made concerning individual demand. Löschian competition does not violate the characteristics when the shape of individual consumer demand functions is altered. Note the following discussion is *not* designed to change the C-V assumptions. Rather, it is presented to challenge one of their main conclusions—the rejection of Löschian competition. Furthermore, the amendment below does not imply agreement that the C-V properties are necessarily the only ones a spatial model should obey, nor does it imply agreement that a spatial model should obey all five of them.[5]

I. Assumptions for Two Models of Löschian Competition

Capozza and Van Order proposed several assumptions which are followed here, including:

ASSUMPTION 1: *A single commodity is produced by all firms and all firms have identical cost functions.*

ASSUMPTION 2: *The cost function is*

$$C=f+cQ \tag{1}$$

where Q is output, C is total production cost, f is fixed cost, and c is marginal cost.

ASSUMPTION 3: *Transport cost per unit of distance is identical between any two points and equal to t dollars per unit.*

ASSUMPTION 4: *Firms can enter freely and will do so until all economic profits disappear.*

Two additional assumptions proposed by Capozza and Van Order will be slightly altered:

ASSUMPTION 5: *Potential consumers occupy a homogeneous unbounded linear space of uniform density D.*

ASSUMPTION 6: *All market areas consist of segments of this linear space.*

Capozza and Van Order chose to work with a plain and with circular markets. The use of linear markets alters the magnitude of the values that are obtained (i.e., profit-maximizing price) *but* the general conclusions derived below continue to hold under the C-V market shape assumptions. Working with the linear market, on the other hand, does substantially simplify the computations which follow.

Löschian behavior characterizes the spatial entrepreneur. That is:

ASSUMPTION 7: *Each entrepreneur assumes his market area is fixed and prices as a monopolist within his space.*

Capozza and Van Order assumed that individual demand was linear, but this need not be the case. Let us start with a general demand function given by

$$P+tu=a-\frac{b}{x}q^{x} \tag{2}$$

where P is mill price, q is demand per consumer, a and b are constants, and u is units of distance.[6] Now x can be varied to encompass a wide range of possible demand functions. For example, if $x=1$, the demand function is linear and the C-V conclusions would follow. If $x>1$, the function is convex upwards. With $x<1$, the function is concave upwards. The function is extremely concave if $x<0$.[7] If $-1<x<0$ and $a=0$, the function describes a constant elasticity demand curve. So x is the vital parameter in determining the shape of the individual demand curve. I must express individual demand in terms of q rather than delivered price $(P+tu)$ in order to obtain the aggregate market demand. This yields

---

[5] All five characteristics are actually intuitively unappealing when put in a spatial context (see my 1979a paper). The characteristics are based upon spaceless competition where alteration of any one parameter effects only supply *or* demand. I have demonstrated elsewhere (1979a) that both costs (supply) *and* demand are impacted when these parameters change in spatial competition.

[6] This general demand function has been used in other spatial pricing models. Several implications of this general function are noted in John Greenhut, Greenhut, and William Kelly, for example, but Löschian competition was not assumed in that paper.

$$q=\left[\frac{x}{b}(a-tu-P)\right]^{1/x} \tag{3}$$

where u can vary from zero to U, the distance to the boundary of a firm's market.

Equation (3) describes an extremely general relationship for individual demand, as x can vary anywhere from -1 to $+\infty$ while still yielding stable equilibrium results. It is possible, however, to depict the general conclusions for any Löschian type spatial firm with two values of x:

ASSUMPTION 8A: $x=1$ *and individual demand is*

$$q=\frac{1}{b}(a-tu-P) \tag{4a}$$

ASSUMPTION 8B: $x=-1/2$ *and individual demand is*

$$q=\left[-\frac{1}{2b}(a-tu-P)\right]^{-2} \tag{4b}$$

We can derive the C-V results with $x=1$ and the derived relationships are representative of all $x>0$ (positive exponential demands). The comparative statics using $x=-1/2$ are the same for all $-1<x<0$ (negative exponential demand).

II. Löschian Reactions under Alternative Demand Conditions

For $x=1$, the aggregate spatial demand is

$$Q_p=D\int_0^U\frac{1}{b}(a-tu-P)\,du \tag{5}$$

$$Q_p=\frac{DU}{b}\left(a-P-\frac{tU}{2}\right) \tag{6}$$

where Q_p represents the aggregation of positive exponential demands. The aggregation of demand with $x=-1/2$ provides

$$Q_n=D\int_0^U\left[-\frac{1}{2b}(a-tu-P)\right]^{-2}du \tag{5$'$}$$

$$Q_n=4b^2D\left(\frac{1}{P-a}-\frac{1}{P-a-tU}\right) \tag{6$'$}$$

where Q_n represents the aggregation of negative exponential demands. Profits are zero in the long run, i.e.,

$$Y=PQ-C=0 \tag{7}$$

Substituting (1) and (6) into (7) yields

$$Y=\frac{DU}{b}\left(a-P-\frac{tU}{2}\right)(P-c)-f=0 \tag{8}$$

while (1) and (6′) into (7) provides

$$Y=4b^2D\left(\frac{1}{P-a}-\frac{1}{P-a-tU}\right)(P-c)-f=0 \tag{8$'$}$$

The alternative Löschian profit-maximizing prices are obtained by maximizing (8) and (8′) with respect to price. Thus

$$P_p=1/2(a+c)-\frac{tU}{4} \tag{9}$$

---

[7] It must be assumed that $x>-1$. If $x<-1$, there will be no profit-maximizing equilibrium, given the assumption of constant marginal cost made earlier. Marginal revenue will be increasing if $x<-1$, and any adjustment will move production away from the $MR=MC$ level.

TABLE 1—COMPARATIVE STATICS FOR LÖSCHIAN COMPETITION UNDER ALTERNATIVE DEMAND CONDITIONS

Parameter Change	Löschian Reaction C-V Results ($x=1$)	$x=-1/2$
Marginal Cost	±	+
Demand Density	+	−
Fixed Costs	−	+
Transport Cost per Unit of Geographic Distance	−	+
Entry of New Firms	+	−

when $x=1$, but

$$(9') \quad P_n = c + \left[(c-a)tU + (c-a)^2\right]^{1/2}$$

for $x=-1/2$.[8, 9]

Let us now examine Löschian competition to see whether it is always inconsistent with the characteristics (C1−C5) which Capozza and Van Order set forth (the comparative statics are summarized in Table 1):

C1) For $x=1$, the profit-maximizing price rises as transportation costs (t) decrease. This contradicts C1, as C-V noted. However, if $x=-1/2$, the profit-maximizing price falls towards marginal cost as transportation costs decline. Characteristic 1 is satisfied only in part, however, by Löschian competition and negative exponential demand. Price does move in the direction that C-V argue it should as parameter t changes, but price does not approach marginal cost as t approaches zero.

Examination of the C-V model reveals perfect competition need not be the structure of their industry even in a spaceless setting. Let $t=0$. Then the cost function (C-V's as well as my equation (1)) results in spaceless imperfect competition or monopoly.[10] The large number of firms required for perfect competition necessitates each producing at a relatively small scale. Thus, price must be greater than marginal cost for survival (unless we assume each of a large number of firms produce and sell an infinite quantity). This long-run cost function has economies of scale (actually, increasing returns) over the entire range so we might expect some firms to expand the scale of their output and lower price. Other firms will therefore be forced out of the industry, and market structure approaches oligopoly or monopoly in this spaceless setting. Free entry prevents profit taking so one firm (or a few firms) may charge a price near marginal cost. Price falls because of scale economies, not because of perfect competition. Furthermore, it is very unlikely that price would equal marginal cost. Also note that as long as $t>0$, a single firm (or a few firms) cannot crowd out all the others. As transport costs fall, profits are available and there is entry. Therefore it is conceivable that spatial competitors operate at smaller and smaller scales as transport costs fall and

---

[8] The behavioral assumption of Löschian competition implies that $\partial U/\partial P=0$ (see C-V). Similar prices were derived in Greenhut, Greenhut, and Kelly.

[9] It should be pointed out that a in equations (4b), (5′), (6′), (8′), and (9′) is not a price intercept as it is in (4a), (5), (6), (8), and (9). As $P\rightarrow\infty$ the function approaches the price axis. The variable a is the asymptote of the curve, or the limit the function approaches as $q\rightarrow\infty$. Also note that in order to have a positive value under the square root $(c-a)>0$.

[10] Another specification of costs has been suggested which results in spatial firms producing efficiently in the long run (see Greenhut, 1971; Ohta, 1976). Justice cannot be done to their interesting arguments here. Briefly however, the spatial entrepreneur has a required return or opportunity cost because he conceives of only a few nearby rivals as his primary competition. As in classical oligopoly theory, he must take the uncertain reaction of these rivals into account when making decisions. It follows that price may exceed marginal production costs (where production costs do not include the subjective required return to the entrepreneur), but production is efficient when the entrepreneur's opportunity costs are taken into account.

market shares shrink. A rise in price should not be surprising when there are economies of scale over the entire range of output.[11] However, the pricing reaction to changing transportation costs is in the direction that Capozza and Van Order argue it should be when $x=-1/2$.

C2) The reactions to changes in fixed costs are also quite different when $x=-1/2$ instead of unity. For $x=1$, price does rise as fixed costs fall. However, if $x=-1/2$, price falls as fixed cost decreases ($\partial P_n/\partial f>0$ in equation (8′)). The C-V directional requirement is fulfilled under this demand condition. The limit requirement is not however, since price does not approach marginal cost as $f\to 0$.

As fixed costs fall, scale economies become less significant. However, as long as fixed costs are positive, one firm (or a few firms) can produce more efficiently than a large number of firms. Thus, in a spaceless setting we would expect imperfect competition or monopoly as long as $f>0$. Of course, an attempt to reap all monopoly profits leads to entry, but it is doubtful that price would approach marginal cost, given $f>0$. In the limit, of course, with $f=0$, free entry implies price equals marginal production cost in a spaceless setting. Thus, this criticism of Löschian competition based on C2 appears to remain valid in part. Even though the comparative statics of Löschian competition with $x=-1/2$ match those desired by Capozza and Van Order, price does not equal marginal cost in the $f=0$ limit. However, I have demonstrated elsewhere (1979b) that price should not equal site-specific marginal production costs as long as $t>0$. Price should exceed plant marginal costs in order to cover society's costs which result from the availability of multiple locations. Multiple locations are desirable, of course, because society's transport costs are reduced with each additional location and effective (or net) demand increases. If price equals site-specific marginal production costs, society's transport costs are relatively high because a relatively small number of locations are available.[12] In other words, an individual firm's costs and revenues do not account for all the social costs and benefits resulting from spatial competition. So price should equal plant marginal production costs only when f *and* t are *both* zero; that is, when space is irrelevant.

C3) As marginal costs rise, the Löschian competitor will raise his price just as C-V argue. Note $\partial P_p/\partial c$ in equation (9), and $\partial P_n/\partial c$ in equation (9′) are obviously both positive.[13]

C4) The Löschian price reaction to a change in density (D) also depends on the shape of the individual demand curve. When demand is linear ($x=1$) price will increase with an increase in density. A higher density permits smaller market areas and the aggregate demand becomes more inelastic when the market area shrinks under an $x=1$. However, when demand is the negative exponential (i.e., $x=-1/2$), price falls as density increases. The aggregate demand becomes more elastic when the extent of the market is reduced in size and $x=-1/2$ (see Greenhut, Greenhut, and Kelly).[14]

C5) As firms enter the industry, the market area shrinks (U becomes smaller). As U becomes smaller, the mill price rises if $x=1$, but falls when x is $-1/2$.

How is it that the shape of the individual demand function can cause such different results in a spatial world? The individual buyer's demand certainly does not play a significant role in spaceless economics. Yet apparently it does in spatial microeconomics.

---

[11] Actually, of course, things are not quite this simple. Shares of existing demand shrink, but as transport costs fall, effective demand rises. Whether firms operate at a larger or smaller scale depends upon the elasticity of firm demand. This point is discussed later. It is also examined in more detail in my 1979a paper.

[12] In fact, free entry and Löschian behavior lead to the socially efficient price and number of plants (see my 1979b paper). Price equals the full marginal cost of production including plant-specific costs and the increase in costs due to multiple locations.

[13] Capozza and Van Order argued that $\partial P_p/\partial c$ is ambiguous. This is true when there is a positive fixed cost (i.e., when we use equation (8)). The relationship is not ambiguous, however, when the assumption of zero fixed cost in the long run is made.

[14] The elasticity of spatial aggregate demand under alternative demand conditions is discussed below, in the context of the relationships of price to density, transport costs, and entry of new firms.

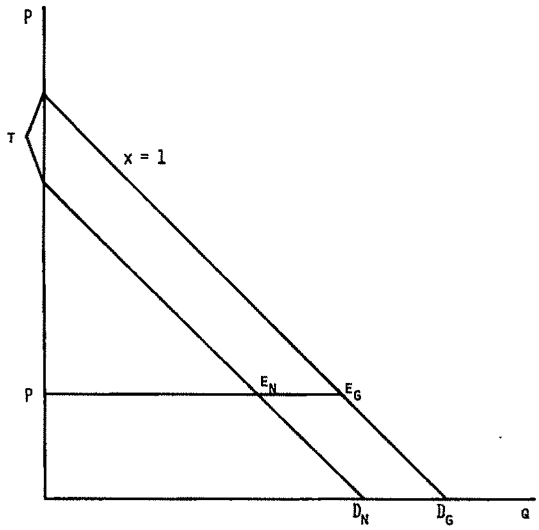


FIGURE 1. NET INDIVIDUAL DEMANDS AND CHANGING ELASTICITY FOR $x = 1$

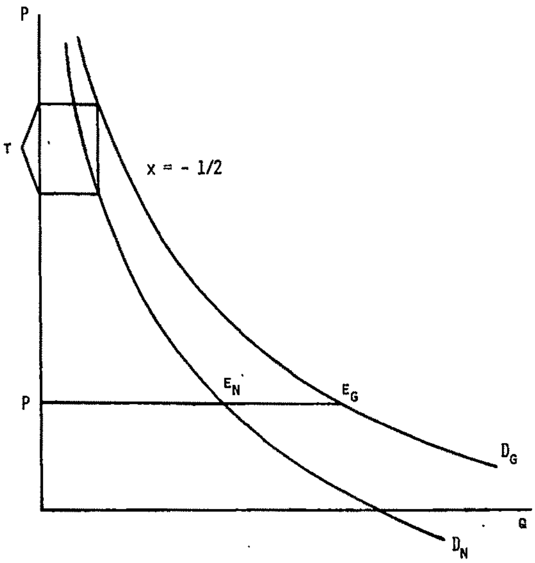


FIGURE 2. NET INDIVIDUAL DEMANDS AND CHANGING ELASTICITY FOR $x = -1/2$

The underlying reason for the different results observed here is the difference between spatial and spaceless aggregate demand. The price received by the firm in spaceless economics equals the price paid by the consumer. Spatial consumers, however, incur buying costs in addition to the price paid to the seller.[15] Producers are therefore subject to demand elasticity that relates to the net market price (net of transportation costs), not the total price paid by consumers (delivered price). The price paid by consumers differs from the price received by producers.

Elasticity at any given price increases when transportation costs are subtracted from a linear gross individual demand curve to obtain net demand (the relevant demand from the seller's point of view). The elasticity of gross demand (D_G) at price P in Figure 1 is less than the elasticity at price P of the demand that is net of transportation costs (D_N) ($E_G < E_N$). These relations apply to any gross demand with $x > 0$. However, the elasticity at a given price decreases as transportation costs are subtracted from any gross demand of the concavity given by $-1 < x < 0$ (as $E_G > E_N$ at price P in Figure 2).

Elasticity of aggregate demand increases at any price as more distant consumers are added and $x > 0$. Elasticity of aggregate demand falls, however, as more distant consumers are added and $-1 < x < 0$, because successively more and more inelastic demands at a given common price are being added to the total.

Anything which causes a spatial firm to lose or gain distant consumers changes the elasticity of the aggregate demand faced by the firm. For example, the impact of entry of a new competitor at a distance is to take away the most elastic demand for the representative seller when $x > 0$. The aggregate demand left for that seller is less elastic than was the original demand. The resulting price effect is upward as Capozza and Van Order noted. However, when $-1 < x < 0$ applies, distant competition removes the least elastic demands from the representative seller's market space. The remaining aggregate demand is then more elastic than was the original aggregate demand. The price effect in this case, contrary to Capozza and Van Order, *is downward.*

---

[15] For a discussion of the difference between spatial and spaceless aggregate demand, see Greenhut (1978).

We can now see more clearly why changes in costs or demand density have the effects noted in Table 1. If transportation costs fall, elasticity of net demand for consumers at each point in the geographic space will decrease when $x>0$. The resulting aggregate demand is less elastic and the price effect is upward. The opposite is true if $-1<x<0$. In that event, the elasticity of individual net demands increases, the aggregate demand becomes more elastic, and the price effect is downward.

A similar argument applies to changes in density. Capozza and Van Order observed that "higher density permits smaller market areas" and "the spatial demand curve is more elastic for smaller market radii" (p. 902). A higher density will lead to short run profits and entry. The market areas of existing firms shrink and the spatial aggregate demand is indeed more inelastic for smaller markets *if* $x>0$. However, if $-1<x<0$, demand becomes more elastic as markets become smaller. Therefore, price rises when $x=1$, but falls when $-1<x<0$.

When fixed costs rise, each firm faces losses. There is exit and the market areas of remaining firms expand. Elasticity of aggregate demand increases as distant buyers are added and $x>0$. The price effect is downward. The opposite is true, however, for $-1<x<0$. Elasticity decreases as distant buyers are added in this case, and the price effect is upward.

Alternative specifications of demand have an impact on theoretical pricing policies because of the effect on the elasticity of *actual* aggregate demand for spatial competitors. All that the alternative behavior assumptions really do is change the elasticity of each firm's *perceived* demand (see Greenhut, Hwang, and Ohta). A Löschian competitor views his demand as more inelastic than does the Hotelling-Smithies entrepreneur that Capozza and Van Order maintain is "...closer to the way most firms behave in most real world situations" (p. 896).[16] When we realize that alternative demand assumptions and alternative behavioral assumptions are just altering the demand elasticities to which a firm reacts, it is not surprising to find that we can obtain similar results by varying either assumption.

III. Conclusions

Each of the C-V "intuitively appealing properties" apply to Löschian competition when demand is negative exponential in form.[17] It follows that one cannot reject Löschian competition for the reasons given by Capozza and Van Order. Perhaps one should reject linear (or positive exponential) demand instead? Of course not. At least three models predict similar results. Capozza and Van Order prefer linear demand and Hotelling-Smithies behavior on the part of entrepreneurs. Greenhut, Hwang, and Ohta proposed positive exponential demand and Greenhut-Ohta behavior. Here we assume negative exponential demand and Löschian competition. To accept one and reject the others requires detailed analysis of all the consequences of each model, (as well as any others that might be proposed) including empirical testing.[18] Until this is effected, one cannot accept any particular assumption concerning entrepreneurial behavior (or individual spatial demand) as being superior to any other assumption.[19] One can only stipulate that there are at least two crucial assumptions in a model of spatial competition, and probably many more.[20]

---

[16] In fact, as Ohta (1980) has noted, the Löschian competitor's perceived demand is his actual demand following entry. The Hotelling-Smithies (and Greenhut-Ohta) demands are perceived demands which are believed to be more elastic than the actual Löschian demand. Thus the Löschian assumption might be restated as, "...firms recognize the actual demand they face and profit maximize accordingly" (my 1979a paper, p. 24). Hotelling-Smithies (and Greenhut-Ohta) competition imply that entrepreneurs are ignorant of their actual demand situation and never learn, even with continued entry and many observations.

[17] Actually, of course, characteristics 1 and 2 are only met in part. The comparative statics are those desired by C-V. However, the limiting requirements (that price approach marginal cost as t or f approach zero) are not met. I maintain that these limiting requirements need not be met in a spaceless world either, however, because the cost function specified by C-V has economies of scale over the entire range of output. Market structure is more likely to be monopoly or oligopoly than perfect competition under such cost conditions. Thus, C-V try to force a spatial model to match the results of spaceless perfect competition when their assumptions would prevent perfect competition in a spaceless setting. For more on this argument see my 1979a paper. Furthermore, I argue elsewhere that price should exceed plant-specific marginal production costs as long as $t>0$ (see my 1979b paper). This results because of additional social costs and benefits that occur in spatial competition which are not accounted for by individual firms. So, even though the limit requirements are not met for characteristics 1 and 2, I maintain that they should not be met.

---

[18] The Greenhut-Ohta assumption of entrepreneurial reactions was not discussed here. This does not mean that C-V's questioning of the Greenhut-Ohta model is accepted. Rather, their model has many desired characteristics which require separate examination.

[19] Neither C-V's preferred model nor the model discussed here will conform completely with spaceless economic conclusions. Inclusion of costly distance still results in many conclusions which contradict classical spaceless economic predictions. We have developed spatial models which conform in some ways with spaceless economics then, but this does not imply that space is not deserving of considerable attention by microeconomists.

[20] For example, both this model and the C-V model assumed constant marginal costs. However, allowing for increasing (and/or decreasing) marginal costs will also affect price reactions (see Greenhut, Greenhut, and Kelly).

REFERENCES

B. Benson, (1979a) "Are the Conclusions of Löschian Spatial Competition Really Counter-Intuitive?," work. paper no. 5-79-1, Pennsylvania State Univ., May 1979.

______, (1979b)"Spatial Competition with Free Entry, Chamberlinian Tangencies, and Social Efficiency," work. paper no. 1-79-1, Pennsylvania State Univ., Sept. 1979.

______, "Spatial Microeconomics: Implications for Market Area Delineation in Anti-Merger Cases," *Antitrust Bull.*, forthcoming.

D. R. Capozza and K. Attaran, "Pricing and Spatial Dispersion of Firms Under Free Entry," *J. Reg. Sci.*, Aug. 1976, *16*, 167–82.

______ **and R. Van Order,** "A Generalized Model of Spatial Competition," *Amer. Econ. Rev.*, Dec. 1978, *68*, 896–907.

J. Greenhut, M. L. Greenhut, and W. H. Kelly, "A Spatial-Theoretical Perspective for Bank Merger Regulations," *Fed. Reserve Bank Chicago Proceedings on Bank Structure and Competition*, 1977, 210–54.

M. L. Greenhut, *A Theory of the Firm in Economic Space*, Austin 1971.

______, "Impacts of Distance on Microeconomic Theory," *Manchester Sch. Econ. Soc. Stud.*, Mar. 1978, *46*, 17–40.

______, **M. Hwang, and H. Ohta,** "Observations on the Shape and Relevance of the Spatial Demand Function," *Econometrica*, July 1975, *43*, 669–82.

J. M. Hartwick, "Lösch's Theorem on Hexagonal Market Areas," *J. Reg. Sci.*, Aug. 1973, *13*, 213–22.

H. Hotelling, "Stability in Competition," *Econ. J.*, Mar. 1929, *39*, 41–57.

A. Lösch, *The Economics of Location*, New York 1954.

E. S. Mills and M. R. Lav, "A Model of Market Areas with Free Entry," *J. Polit. Econ.*, June 1964, *72*, 278–88.

H. Ohta, "On Efficiency of Production Under Conditions of Imperfect Competition," *Southern Econ. J.*, Oct. 1976, *43*, 1124–35.

______, "Spatial Competition, Concentration, and Welfare," *Reg. Sci. Urban Econ.*, Mar. 1980, *10*, 3–16.

A. Smithies, "Optimum Location in Spatial Competition," *J. Polit. Econ.*, Jan. 1941, *49*, 423–39.

N. H. Stern, "The Optimal Size of Market Areas," *J. Econ. Theory*, Apr. 1972, *4*, 154–73.

NOTES

The ninety-fourth annual meeting of the American Economic Association will be held in Washington, DC, December 28–30, 1981. The Professional Placement Service will be open December 27–30.

Call for Papers for the 1981 Meetings

Members wishing to give papers or make suggestions for the program for the meetings to be held in Washington, December 28–30, 1981, are invited to send their ideas to Professor Gardner Ackley, Economics Department, University of Michigan, Ann Arbor, MI 48109. As in the past, the sessions sponsored by the American Economic Association will be of two types: invited papers and contributed papers. Most of the invited papers will be published in the *Papers and Proceedings* issue to appear in May 1982; the contributed papers will not be so published. Suggestions of topics or proposals of papers for the invited sessions should be submitted as soon as possible. Abstracts of proposed contributed papers must be received no later than March 1, 1981. Econometric studies or highly mathematical papers are not appropriate for the sessions sponsored by the American Economic Association. However, members wishing to present econometric papers may submit their proposals or abstracts to the Econometric Society which meets with the American Economic Association, and normally schedules a number of sessions involving such papers.

The Division of Information Science and Technology of the National Science Foundation supports studies of the role of information in the economy. General topics include but are not limited to: modelling of information flow at various levels of aggregation; development of relevant measures and empirical techniques, studies of the comparative complementarity and substitutability of information versus noninformation inputs; characterization of scientific and technical information as an input in production and as a product; economies of scale in the production, dissemination and use of information, and the resulting market structures; and the joint use of information by firms, industries and nations, and the properties of the resulting structural arrangements. Proposals may be submitted at any time. Approximately six months is needed for review and processing of proposals. Additional information may be obtained by writing to the Division of Information Science and Technology, National Science Foundation, 1800 G Street, NW, Washington, DC 20550 or by calling the Division at 202+357-9569.

The Postal Rate Commission, an independent federal regulatory agency, announces the initiation of a doctoral dissertation program designed to encourage doctoral candidates of varied academic disciplines to direct their studies toward the analysis of postal issues and problems. The Rate Commission will provide funds to doctoral candidates at accredited institutions of higher learning who desire to pursue uninterrupted, full-time research in various areas of interest related to the U.S. postal system. For further information, write Doctoral Dissertation Grant Board, Office of Technical Analysis and Planning, Postal Rate Commission, 2000 L Street, NW, Suite 500, Washington, DC 20268.

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to air fare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for applications to be received in the ACLS office are: meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 345 East 46th St., New York, NY 10017, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting. Even when plans are incomplete, a prospective applicant should request forms in advance of the cut-off date, since deadlines are firm and no exceptions are permitted. Awards will be announced approximately two months after each deadline.

The Visiting Lecturer Program in Statistics is continuing into its eighteenth successive year, and is available to American and Canadian schools. The program is sponsored jointly by the American Statistical Association, the Biometric Society, and the Institute of Mathematical Statistics. Leading teachers and research workers in statistics from universities, industry, and government participate as lecturers. Topics include subjects in experimental and theoretical statistics, as well as in such related areas as probability theory, information theory, and stochastic models in the physical, biological, and social sciences. The purpose of the program is to provide information to students and college faculty about the nature and scope of modern statistics, and to provide advice about careers, graduate study, and college curricula in statistics. Inquiries should be addressed to Barry C. Arnold, Chair, Visiting Lecture Program in Statistics, Department of Statistics, University of California, Riverside, CA 92521.

The ninth annual Telecommunications Policy Research Conference will be held April 26–29, 1981. It will provide a forum for analysis and discussion of important telecommunications policy issues. Participants will include researchers and policymakers from academia, government, and industry. Authors of papers selected for presentation will be reimbursed for travel and conference living expenses if no alternative source of funding is available. Send proposals and inquiries to TPRC Organizing Committee, Att. William E. Taylor, Bell Laboratories 2C-258, 600 Mountain Avenue, Murray Hill, New Jersey 07974.

The P. W. S. Andrews Memorial Prize is awarded annually for an essay by a young scholar (under the age of 30 or within eight years of taking his/her first degree) in the general field of Industrial Economics and the Theory of the Firm, broadly interpreted. The prize is £300 (or the equivalent in other currency) and the winning essay will normally be published in *The Journal of Industrial Economics*. Submissions should be original research by the candidate only, not previously published, and not previously awarded any other prize. They should be in English, and should not normally exceed 10,000 words. The closing date for entries is December 31 in each year. Details may be obtained from the Administrative Officer, Office for Student and College Affairs, University House, University of Lancaster, Lancaster, LA1 4YW, England.

The third Conference on Major International Economic Issues will be held at the University of Southern California, June 13-14, 1981. The theme is Multinational Corporations in a Growing World Economy. Submit proposals and inquiries to Professor Nake M. Kamrany, Department of Economics, University of Southern California, Los Angeles, CA 90007.

The annual meeting of the Rocky Mountain Association for Slavic Studies will be held in conjunction with the Western Social Science Association meeting, April 23-25, 1981, in San Diego, California. It will focus on Slavic studies (the *USSR* and Eastern Europe) and communist studies (no geographical limitation). Submit proposals and inquiries to Don Chenoweth, Department of Social Science, Cameron University, Lawton, Oklahoma 73505.

The Population Reference Bureau announces a new Visiting Scholar Program for demographers and other population specialists. An annual fellowship will be awarded to a qualified mid-career demographer/population specialist. The person selected will be stationed at the Washington office of the Bureau for one academic year or, in certain circumstances, one semester. He or she will be expected to develop and complete a specific demographic research project resulting in one or more *PRB* publications dealing at least in part with policy implications, domestic or international. Interested persons should contact the Committee on the Visiting Scholar Program, PRB, 1337 Connecticut Avenue, NW, Washington, DC 20036 (telephone 202+785-4664).

The theme of the December 1981 issue of the *Review of Social Economy* will be "John Maurice Clark—Social Economist." Papers are cordially invited and should be sent to the Editor, *RSE*, DePaul University, 25 E. Jackson Blvd, Chicago, Illinois 60604 no later than February 15, 1981. A fee of $20 is required to partially cover the cost of processing; members of the Association for Social Economics are entitled to submit one manuscript a year without charge.

The fourth International Time-Series Meeting (ITSM) will be held in Valencia, Spain, June 22-26, 1981. For further information, contact Oliver Anderson, ITSM Spain, 9 Ingham Grove, Lenton Gardens, Nottingham NG7 2LQ, England.

The Southern Regional Science Association will hold its next meetings in Washington, DC, April 14-16, 1981. For further information, contact Dr. Richard Olsen, Program Chairman, Applied Business Research, Inc., 20 LaFayette Circle, Wellesley, MA 02181.

The S. S. Huebner Foundation for Insurance Education, a nonprofit foundation affiliated with the University of Pennsylvania, is sponsoring its sixth annual research grants competition. The grants are intended to support research in the field of risk and insurance. Full-time faculty members at colleges and universities in the United States and Canada are eligible to apply for grants. Applicants should hold a terminal degree such as Ph.D. or D.B.A., a law degree, or be a fellow of an actuarial society. Among the general topic areas which will be considered for grants are risk theory, consumer demand for insurance, risk management, health insurance, other methods of health care financing, social insurance programs, international insurance issues, insurance law, and insurance regulation. Grants are available from the Foundation in amounts up to $10,000. Proposals must be submitted by March 1, 1981, and the grants will be awarded by June 1, 1981. Additional information regarding the program can be obtained by writing to Dr. J. David Cummins, Associate Director, S. S. Huebner Foundation for Insurance Education, Colonial Penn Center, 3641 Locust Walk CE, Philadelphia, PA 19104.

The *Economic Forum* is pleased to announce the third annual Rasmussen Prize Essay in Political Economy. The winner of this award will receive $500 and publication in the Summer 1981 issue of the *Economic Forum*. The essay competition is open to all students who submit manuscripts on any area in economics. The submission deadline is April 1, 1981. Ramchandran Jaikumar (The Wharton School) and Edward M. Ochoa (The New School for Social Research) were chosen joint winners of the second Rasmussen Prize Essay competition; their articles and an announcement of the essay competition appear in the Summer 1980 issue of the *Economic Forum*. For additional information, contact Editor, *Economic Forum*, University of Utah, Salt Lake City, Utah 84112.

Omicron Delta Epsilon, The International Honor Society in Economics, invites the sumbission of entries for the thirteenth annual Irving Fisher Graduate Monograph and Frank W. Taussig Award Undergraduate Competitions. The Fisher Award consists of $1,000 and publication as a book by Princeton University Press, subject to approval of its editorial board. In addition, the winner will be invited to submit a paper based on the winning entry to the *American Economic Review*. The recommendations of the Final Selection Board of the Competition will be considered by the *Review* in the refereeing process. All finalists will be invited to submit a paper for publication in *The American Economist*. The Taussig Award consists of $250 to the winner, $100 to his/her chapter, and publication of the paper in *The American Economist*. Entries for the Fisher Award should be submitted to Departmental Selection Committees by January 1, 1981 and entries for the Taussig Award by May 15, 1981. They will be judged by the International Editorial Board and finalists by the Final Selection Board, consisting of Professors John S. Chipman, Carl F. Christ, James Mirrlees, Arnold Zellner, and Egon Neuberger (editor). Anyone interested in entering the competitions should contact E. Neuberger, Editor, Department of Economics, State University of New York, Stony Brook, NY 11794, for information on entrance requirements.

The John M. Olin Fellowship Program in Law and Economics, administered by the Law and Economics Center at Emory University, is accepting applications for the class entering in August 1981. The three-year program leads to the Doctor of Law degree and is designed for highly motivated individuals with a strong foundation in graduate microeconomics and its application. The course of study prepares fellows for scholarly work in various fields of law and economics and opens new career oportunities in the law. An increasing proportion of successful candidates have the Ph.D. and have been teaching. Individuals who will have passed the preliminary examinations for the Ph.D. by August 1981 are also encouraged to apply and to plan to complete their dissertation while in residence at Emory University. Five fellowships are available each year. These provide full tuition and fees (currently about $4,600 per year) plus stipends of at least $8,000 per year ($7,000 per year for fellows without the Ph.D.) Up to five additional Fellowships of lesser amounts may also be offered each year. Deadline for submitting applications is February 1, 1981. Awards are announced by April 1, 1981. Candidates must apply separately to the School of Law and should take the Law School Admissions Test no later than December 1980. For more information write to John M. Olin Fellowship Program, Law and Economics Center, Emory University School of Law, Atalnta, GA 30322.

The Carolina Population Center invites applications for a small number of postdoctoral fellowships for 1981–82 for population research training in the following fields: sociology, psychology, economics, anthropology, epidemiology, and maternal and child health. Postdoctoral programs of the following types will be considered: (1) Specialized training in the discipline of the applicant's doctorate for those holding doctorates involving population research training; (2) Training in a second discipline for those with doctorates involving population research training; (3) Individualized interdisciplinary programs for those with doctorates which may (but need not) involve population research training. Applicants must be U.S. citizens or permanent residents of the United States. The application deadline is February 1, 1981. Direct inquiries to J. Richard Udry, Director, Carolina Population Center, University of North Carolina, Chapel Hill, NC 27514.

The Center for Asian Economic Research is sponsoring a conference on April 16–17, 1981, at the New Brunswick campus of the State University of New Jersey. The theme is "Asian-American Economic Relations." Invitees from Asian countries and the United States will confer on topics such as: trade; trade barriers; investments–private and public; multinationals–Asian and American; economic growth; income redistribution; social change. Please address all communications to Professor M. Dutta, Director, Center for Asian Economic Research, Rutgers University, 180 College Avenue, New Brunswick, NJ 08903.

The fifth biannual Arne Ryde Symposium, "The Economics of Social Insurance," is being arranged by the Department of Economics at the University of Lund, to be held in September 1981. The focus will be on theories, policies, and institutions concerning social insurance. Address all inquiries to Carl Hampus Lyttkens, Department of Economics, Box 5137, S-220 05 Lund, Sweden (telephone 046/108652).

Deaths

James R. Nelson, professor of economics, Amherst College, Apr. 30, 1980.

Raymond P. Powell, professor of economics and Russian studies, Yale University, May 18, 1980.

Retirements

William Carroll, department of economics, Drew University.

Charles Drake, Central Missouri State University: visiting professor, department of economics, Southwest Missouri State University, 1980-81.

Robert Triffin, professor, department of economics, Yale University, June 1980.

Visiting Foreign Scholars

Gideon Fishelson, Tel-Aviv University: visiting associate professor, department of economics, Brown University.

Gerhard Mensch, International Institute of Management, Berlin: visiting professor of economics and management, Case Western Reserve University, Sept. 1980–May 1981.

Bjorn Wahlroos, Swedish School of Economics and Business Administration: visiting assistant professor, department of economics, Brown University, Sept. 1980–June 1981.

Promotions

Ashok Bhargava: professor of economics, University of Wisconsin-Whitewater, June 1980.

George Borjas: associate professor, department of economics, University of California-Santa Barbara, July 1, 1980.

John J. Casson: vice president, corporate strategy and chief economist, American Express Company, Apr. 30, 1980.

Donald P. Cole: professor of economics, Drew University.

Robert M. Feinberg: associate professor of economics, Pennsylvania State University, July 1, 1980.

Morley K. Gunderson: professor of economics, University of Toronto, July 1, 1980.

Herbert Kaufman: professor of economics, Arizona State University, Aug. 20, 1980.

Stephen LeRoy: associate professor, department of economics, University of California-Santa Barbara, July 1, 1980.

Lee McPheters: professor of economics, Arizona State University, Aug. 20, 1980.

G. Franklin Mathewson: professor of economics, University of Toronto, July 1, 1980.

Yash P. Mehra: associate professor of economics, Pennsylvania State University, July 1, 1980.

Marc A. Miles: associate professor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Norman Mintz: senior vice president, Columbia University, July 1, 1980.

Thomas M. Stevenson: professor of economics, St. Louis University, Sept. 1, 1980.

Brian D. Wright: associate professor, department of economics, Yale University, July 1, 1980.

Administrative Appointments

Abby J. Cohen: vice president, T. Rowe Price Associates, Inc., Baltimore, Apr. 21, 1980.

William B. Neenan, University of Michigan: dean, College of Arts and Sciences, Boston College, June 2, 1980.

Walter Nicholson: chairman, department of economics, Amherst College, July 1, 1980.

Rama V. Ramachandran: chairman, department of economics, Southern Methodist University, July 1980.

Sidney I. Simon: chairperson, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Irving J. Spitzberg, Jr., State University of New York-Buffalo: general secretary, American Association of University Professors, July 1, 1980.

Patrick J. Welch: chairman, department of economics, St. Louis University, July 1, 1980.

Appointments

Leslie A. Austin: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Seymour D. Barcun: lecturer, department of economics, Rutgers College-New Brunswick, July 1, 1980.

James T. Barrese: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

David B. Crary: assistant professor, department of economics, Eastern Michigan University, Sept. 1, 1980.

Fred Curtis: instructor, department of economics, Drew University.

Allen Dalton, Virginia Polytechnic Institute and State University: assistant professor of economics, Arizona State University, Aug. 20, 1980.

Gregory L. Dow, University of Michigan: assistant professor, department of economics, Yale University, July 1980.

Zvi Eckstein, University of Minnesota: assistant professor, department of economics, Yale University, July 1980.

Susan F. Feiner: visiting assistant professor, department of economics, College of William and Mary, Sept. 1980.

Ronald E. Grieson, Princeton University: professor of economics, University of California-Santa Cruz, July 1, 1980.

Louis H. Henry: visiting associate professor, department of economics, College of William and Mary, Sept. 1980.

Paul D. Holtberg: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Russell A. Janis: assistant professor of economics, Amherst College, July 1, 1980.

Christiana Kelton, University of Wisconsin: assistant professor of economics, Case Western Reserve University, Sept. 1980.

John Kennan, Brown University: department of economics, University of Iowa.

Chongyoul Kim: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Gary D. Koppenhaver, University of Iowa: assistant professor, department of economics, Southern Methodist University, Fall 1980.

Laurence J. Kotlikoff, University of California-Los Angeles: assistant professor, department of economics, Yale University, July 1980.

Kasibhatla Krishnamoorti: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

David T. Levy: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Marc Lieberman, Princeton University: assistant professor of economics, University of California-Santa Cruz, July 1, 1980.

Anuradha Luther, Stanford University: assistant professor of economics, University of California-Santa Cruz, July 1, 1980.

Gordon McMeekin, GE TEMPO: assistant professor of economics, Case Western Reserve University, June 1980.

Michael Melvin, University of California-Los Angeles: assistant professor of economics, Arizona State University, Aug. 20, 1980.

Jose Mendez, Southern Methodist University: assistant professor of economics, Arizona State University, Aug. 20, 1980.

Ezra J. Mishan: professor of economics, Colgate University, fall 1980.

Peter J. Montiel, Florida International University: visiting assistant professor, Amherst College, July 1, 1980.

Kemper Moreland: assistant professor, economics department, Eastern Michigan University, Sept. 1, 1980.

Peggy B. Musgrave, Vassar College: professor of economics, University of California-Santa Cruz, July 1, 1980.

Richard A. Musgrave, Harvard University: adjunct professor of economics, University of California-Santa Cruz, July 1, 1980.

William B. Neenan, University of Michigan: professor of economics, Boston College, June 2, 1980.

Hiam Ofek: associate professor of economics, Colgate University, 1980-81.

Charles Ofori-Mensa: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Milton Pappas, Johns Hopkins University: department of economics, Pennsylvania State University, Sept. 1, 1980.

William C. Perkins: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Susan Pozo, Michigan State University: department of economics, Pennsylvania State University, Sept. 1, 1980.

Louis Putterman, Yale University: assistant professor, department of economics, Brown University.

William J. Reid, Jr.: visiting assistant professor, department of economics, College of William and Mary, Sept. 1980.

Mark J. Roberts, University of Wisconsin: department of economics, Pennsylvania State University, Sept. 1, 1980.

Scott Saunders: assistant professor of economics, Colgate University.

Barry Schachter: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

John Schroeter, University of Minnesota: assistant professor of economics, Arizona State University, Aug. 20, 1980.

Rosalind Seneca: associate professor, department of economics, Drew University.

Douglas R. Shaller: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

David Shapiro, Ohio State University: department of economics, Pennsylvania State University, Sept. 1, 1980.

Joel Sobel: assistant professor, department of economics, University of California-San Diego, July 1, 1980.

James B. Stewart, University of Notre Dame: department of economics, Pennsylvania State University, Sept. 1, 1980.

Roswell G. Townsend, Wilson College: professor of economics, Chatam College.

Hiroki Tsurumi, Rutgers University: visiting professor, department of economics, Pennsylvania State University, Sept. 1, 1980.

Eugene N. White: instructor, department of economics, Rutgers College-New Brunswick, July 1, 1980.

Halbert White, University of Rochester: associate professor, department of economics, University of California-San Diego, July 1, 1980.

Willard E. Witte, Pennsylvania State University: visiting assistant professor, Indiana University, July 1, 1980–June 30, 1981.

Leaves for Special Appointment

Ralph E. Beals, Amherst College: Harvard Institute for International Development, adviser to the Ministry of Finance, Government of Indonesia, 1980-82.

Peter J. Coughlin, Stanford University: postdoctoral research fellow, Oxford University, 1980-81.

Michael Duffy, Arizona State University: Center for Naval Analysis.

Michael Greenwood, Arizona State University: University of Colorado, Aug. 20, 1980.

William Poole, Brown University: Reserve Bank of Australia, Sydney, July 1, 1980–July 1, 1981.

George Provenzano, University of Illinois: Office of Strategic Assessments and Special Studies, U.S. Environmental Protection Agency, 1980-81.

Michael J. Wasylenko, Pennsylvania State University: U.S. Department of Housing, June 1980–June 1981.

Nancy Wentzler, Pennsylvania State University: Brookings Institution Fellowship Program, Sept. 1, 1980–Aug. 31, 1981.

Resignations

Mary Ann Baily, Yale University: staff economist, President's Commission, Study of Ethics in Medicine and Biomedical and Behavioral Research, July 1980.

Richard N. Cooper, Yale University: Harvard University, Dec. 31, 1979.

Sylvester R. Damus, University of Winnipeg: Economic Council of Canada, Ottawa, Sept. 1, 1980.

Klaus Friedrich, Pennsylvania State University: economist, Division of International Finance, Board of Governors of the Federal Reserve System, July 1, 1980.

Paul R. Krugman, Yale University: associate professor, Massachusetts Institute of Technology, July 1980.

Thomas Layman, Arizona State University: Crocker Bank, San Francisco, Aug. 1, 1980.

John M. Quigley, Yale Univeristy: professor of public policy, University of California-Berkeley, June 1980.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Editorial Office, University of California, Los Angeles, CA 90024.

◆

Effective January 1, 1981, Association membership dues for 1981, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows: $30.00 for regular members with rank of assistant professor or lower, or with annual income of $14,400 or less; $36.00 for regular members with rank of associate professor, or with annual income of $14,400 to $24,000; $42.00 for regular members with rank of full professor, or with annual income above $24,000; $15.00 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $100.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee. In countries other than the United States, add $5.00 to cover extra postage.

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1981 issue of the *Review* will carry the seventy-eighth list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1981. This announcement is an invitation to send us information for the preparation of the list. This announcement supersedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1981 issue the names of those who will have been awarded the doctoral degree since June 1980 and the titles of their dissertations. Dissertation abstracts will no longer be published, as these are published elsewhere.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ________

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ________________________

Degree Conferred (Ph.D. or D.B.A.) ____________ Year ____________

Dissertation Title: ________________________

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Editorial Office, University of California, Los Angeles, CA 90024.

◆

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed, and published by the American Economic Association, a nonprofit educational organization, located at 1313 21st Avenue So., Nashville, Davidson County, Tennessee 37212. The Managing Editor is Professor George H. Borts, *American Economic Review*, Brown University, Robinson Hall, Providence, RI 02912. During the preceding 12 months the average number of copies printed for each issue was 27,539, the average paid circulation, 24,479; the average free distribution, 400; the average number of copies distributed, 24,879. Corresponding figures for the last issue before filing: 27,500 total number copies printed; 24,136 total paid circulation; 400 free copies distributed; 24,536 total distribution.

SEVENTY-SEVENTH LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1980

General Economics; including Economic Theory History of Thought, Methodology, Economic History, and Economic Systems

LETTY D. ANDERSON, Ph.D. Northwestern 1980. The diffusion of technology in the nineteenth-century American city: Municipal water supply investments.

KAZUHIRO ARAI, Ph.D. Illinois 1979. Expected cost-minimizing search and expected utility-maximizing search.

YVES M. BALCER, Ph.D. Massachusetts Institute of Technology 1980. Four essays in economics of uncertainty.

PAULINA BEATO, Ph.D. Minnesota 1979. Marginal cost pricing and increasing returns.

BIMAL K. BERI, Ph.D. Johns Hopkins 1980. A study of problems associated with transfer of wealth under conditions of uncertainty.

GAUTAM BHATTACHARYA, Ph.D. Rochester 1980. Two essays in economics of uncertainty: The importance of learning in planning with exhaustible resources and optimality of equilibrium with incomplete markets.

FORREST C. BLODGETT, Ph.D. Portland State 1979. Institutional economics and urban political economy.

KIM C. BORDER, Ph.D. Minnesota 1979. Social welfare functions and large economies.

RANDALL L. CALVERT, Ph.D. California Institute of Technology 1980. On the role of imperfect information in electoral politics.

ANN M. CARLOS, Ph.D. Western Ontario 1980. The North American fur trade, 1804-21: A study in the life cycle of a duopoly.

JEFFREY CARMICHAEL, Ph.D. Princeton 1979. The role of government financial policy in economic growth.

THOMAS H. CATE, Ph.D. Florida State 1979. The intellectual origins of John Maynard Keynes' policy recommendations.

TERRY CHACON, Ph.D. California (San Diego) 1979. Public employment in a macro-economic model.

SHEETAL K. CHAND, Ph.D. Western Ontario 1980. Inflation, unemployment, and neoclassical monetary equilibrium.

HENRY W. CHAPPELL, Ph.D. Yale 1979. The economics of campaign finance: An empirical study.

ADHIP CHAUDHURI, Ph.D. Columbia 1980. Envy, distributive justice, and social choice.

BARRY S. CLARK, Ph.D. Massachusetts 1980. Justice and economic theory.

RICHARD N. CLARKE, Ph.D. Harvard 1980. Differential game models of oligopoly: Static and dynamic; deterministic and stochastic.

SWAPAN K. DAS GUPTA, Ph.D. Rochester 1980. Three essays on intertemporal economics.

SAMAR K. DATTA, Ph.D. Rochester 1980. Choice of agricultural tenancy in the presence of uncertainty and transaction costs.

TERENCE M. DWYER, Ph.D. Harvard 1980. A history of the theory of land value taxation.

PHILIP H. DYBVIG, Ph.D. Yale 1979. Recovering additive utility functions.

TOSHITAKA FUKIHARU, Ph.D. Rochester 1980. On the stability of the competitive economy.

JOHN D. GEANAKOPLOS, Ph.D. Harvard 1980. Four essays on the model of Arrow and Debreu.

JOEL C. GIBBONS, Ph.D. Chicago 1979. Flexibility in production and the theory of fixed capital.

GENE D. GUILL, Ph.D. Duke 1979. Structural change in the Soviet Union.

MICHAEL J. HALEY, Ph.D. Pennsylvania 1979. A dynamic theory of a firm that learns by doing.

FUMIO HAYASHI, Ph.D. Harvard 1980. Rational expectations and the consumption function.

ANNE R. HORNSBY, Ph.D. Georgia State 1980. Shifts in the distribution of wealth among blacks in Georgia, 1890-1915.

JAVIER IGUINIZ, Ph.D. New School 1980. Elements towards a more general theory of unequal exchange.

LYLE D. ISRAELSEN, Ph.D. Massachusetts Institute of Technology 1980. The determinants of Russian state income, 1800-1914: An econometric analysis.

ANTHONY JACKSON, Ph.D. Minnesota 1979. A dominance approach to optimal taxation.

MARK S. KAMLET, Ph.D. California (Berkeley) 1980. Resource allocation in market-insulated bureaucracies: Two case studies.

ARNOLD S. KLING, Ph.D. Massachusetts Institute of Technology 1980. The cost to a firm of changing prices: A theory and its implications.

LEWIS A. KORNHAUSER, Ph.D. California (Berkeley) 1980. Breach of contract in the presence of risk aversion, reliance, and reputation.

CHARLES R. LAINE, Ph.D. California (Davis) 1980. Selecting sets of alternatives: A comparison of voting schemes.

WILLIAM R. LASHER, Ph.D. Southern Methodist 1979. Dynamic oligopolistic pricing strategies with learning.

JOHN LATHROP, Ph.D. Michigan 1979. Preferences for consumption streams over time: Some experimental results.

DAVID B. LAWRENCE, Ph.D. Iowa State 1979. The quantification of the value of information in decision making.

MARK J. MACHINA, Ph.D. Massachusetts Institute of Technology 1980. Essays in the economics of uncertainty.

RODNEY R. MADDOCK, Ph.D. Duke 1979. Rational expectations, political business cycles, and the course of macro-economic theory.

GEOFFREY T. MILLS, Ph.D. Illinois 1979. The economics of price control: The OPA experience, 1941-46.

CRAIG L. MOSER, Ph.D. Ohio State 1979. An economic analysis of the Soviet educational system, 1945-75.

CLARK NARDINELLI, Ph.D. Chicago 1980. An economic history of the Factory Acts.

RAMON NEMESIO, Ph.D. Harvard 1979. Personal development, democracy in production organizations, and the socioeconomic system.

BRENDAN O'FLAHERTY, Ph.D. Harvard 1980. Monopoly, local government, and distribution.

TIMOTHY J. O'NEILL, Ph.D. Duke 1979. The role of private, nonprofit research organizations in policy formation: A case study of the Atlantic Provinces Economic Council and the Institute of Public Affairs.

MARTIN J. OSBORNE, Ph.D. Stanford 1979. An analysis of power in exchange economies.

WILLIAM H. PHILLIPS, JR., Ph.D. Massachusetts Institute of Technology 1980. The economic performance of late Victorian Britain: Traditional historians and growth.

LAWRENCE B. PULLEY, Ph.D. Virginia 1980. Mean variance approximations to the growth-optimal decision rule: An empirical investigation.

POONG RHEE, Ph.D. Illinois 1979. An analysis of macro-economic disequilibrium in the Soviet Union.

ROBERT J. ROSSANA, Ph.D. Johns Hopkins 1980. Inventory investment and the short-run demand for labor.

JOHN ROWCROFT, Ph.D. Simon Fraser 1979. The neoclassical theory of the firm.

JOSEPH SALERNO, Ph.D. Rutgers 1980. The doctrinal antecedents of the monetary approach to the balance of payments.

GARRY J. SCHINASI, Ph.D. Columbia 1980. A non-linear dynamic disequilibrium model of macro-economic fluctuations: A theoretical and empirical study.

GERALD A. SMITH, Ph.D. Louisiana State 1979. A historical and methodological study of means and ends in economic theory.

CHESTER S. SPATT, Ph.D. Pennsylvania 1979. Three essays on the economics of finance and organization.

PADMANABHAN SRINAGESH, Ph.D. Rochester 1980. Welfare analysis of non-linear prices.

GERALDINE L. SZYMANSKI, Ph.D. Northwestern 1980. Uncertainty, portfolio allocation, and the determinants of consumer spending.

MOHAMMAD TABIBIAN, Ph.D. Duke 1980. An investigation in the theory of consumer behavior.

FUJIO TANAKA, Ph.D. New School 1980. The reproduction schemes of Marx and Keynes.

JOHN TRESSLER, Ph.D. Missouri (Columbia) 1980. The theory of the firm and industry in a competitive economy with demand uncertainty.

RANJIT VOHRA, Ph.D. Connecticut 1980. Three essays on optimal endogenous technical progress.

STEVEN S. WILDMAN, Ph.D. Stanford 1979. Advertising, consumer learning, and competitive strategies.

GREGORY R. WOIROL, Ph.D. California (Berkeley) 1980. Economics as an empirical science: A case study.

MICHAEL L. WYZAN, Ph.D. North Carolina (Chapel Hill) 1979. A production function approach to the estimation of opportunity cost prices for five Soviet agricultural commodities.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

OSMAN S. AHMED, Ph.D. North Carolina (Chapel Hill) 1979. The potential effects of income redistribution on output, employment, imports, savings, and the pattern of income distribution: The case of Kenya.

SOWAH ANYETEI, Ph.D. Northwestern 1980. An econometric analysis of macro-economic policy in Ghana, 1956-69.

HOSSEIN L. ARJOMAND, Ph.D. Oklahoma 1980. The side effects of Public Law 480 and implications for economic development: A case study of Iran.

HASHIM A-SHAMI, Ph.D. Wisconsin (Madison) 1979. The impact of cotton on economic development of Egypt, 1952-76.

MILTON P. ASSIS, Ph.D. Johns Hopkins 1980. A short-run macro-economic policy model for Brazil.

HYUNG JONG BAEK, Ph.D. Columbia 1980. Economic growth and income inequality: A theoretical analysis.

DANIEL O. BOAMAH, Ph.D. Western Ontario 1980. Towards an optimal cocoa policy in Ghana: A dynamic optimization approach.

ANTONIO M. C. G. BORGES, Ph.D. Stanford 1980. Two general equilibrium models of the U.S. economy.

PETER BRUST, Ph.D. Indiana 1980. Inventory's role in an open economy's dynamic adjustment.

EUFRONIO CARRENO, Ph.D. Rutgers 1980. The Brazilian model of development.

MICHAEL G. CARTER, Ph.D. Stanford 1979. Political business cycles in the United States: Theory and evidence.

Andre Chaib, Ph.D. Michigan 1979. The export performance of a small, open developing economy: The Lebanese experience, 1951-74.

Klin-Keo Chintakananda, Ph.D. Indiana 1980. The role of investment in education in Thailand's economic and social development, 1961-76.

William H. Courtney, Ph.D. Brown 1980. Manufacturing production functions for industries and multinational corporations in less-developed countries.

Celso R. Crocomo, Ph.D. Michigan State 1979. Risk-efficient fertilizer rates: An application to corn production in the Cerrado region of Brazil.

Javier S. M. del Campo, Ph.D. Duke 1979. A rational expectations macro-econometric model of the Mexican economy.

Douglas P. Dungan, Ph.D. Princeton 1980. An empirical, multisectoral, Walrasian-Keynesian model of the Canadian economy.

George W. Evans III, Ph.D. California (Berkeley) 1980. Inflation dispersion in the post-World War II United States: A study of the relationship between aggregate and relative price change.

Kenneth S. Flamm, Ph.D. Massachusetts Institute of Technology 1980. Technology, employment, and direct foreign investment: Evidence from the Mexican manufacturing sector.

J.R. Deep Ford, Ph.D. Purdue 1979. Domestic resource costs and development policy: An analysis of Guyana's milk supply sector.

Jeffrey S. Hammer, Ph.D. Massachusetts Institute of Technology 1980. Essays in economic development and income distribution.

Nancy L. Happe, Ph.D. Princeton 1980. South Korea in the wake of the oil price rise: The economics of a large exogenous shock.

Malcolm Harris, Ph.D. Connecticut 1980. A model of inventory investment during recession, with special reference to the 1973-75 downturn.

Julia Herszkowicz, Ph.D. Purdue 1979. Resource allocation under risk in a mixed economy.

Barbara L. Hiller, Ph.D. California (Berkeley) 1979. External debt and economic development: The debt crisis in Peru.

Frederick I. Johnson, Ph.D. Minnesota 1980. An analysis of Brazil's sugar program, 1953-75.

Woo Sik Jung, Ph.D. California (Berkeley) 1979. Expectations and optimal stabilization policies.

Ka Yanape L. Kamoto, Ph.D. Pittsburgh 1980. Investment and investment policy in a small, open economy: The Zairean case, 1965-75.

Pichai Kanivichaporn, Ph.D. Pennsylvania 1979. An econometric study of the Asian rice economies.

Homi J. Kharas, Ph.D. Harvard 1980. Optimal debt accumulation for less-developed countries: Theory and practice.

James Kilpatrick, Ph.D. Michigan 1979. The use of fertilizer in China's agricultural development.

Chang-Joon Kim, Ph.D. Cincinnati 1980. A study of money and bank credit as target variables in the Korean economy.

Jae Won Kim, Ph.D. Ohio State 1980. Capital imports, employment and economic growth in a dualistic open economy: The case of Korea.

Kil-Joong Kim, Ph.D. Cincinnati 1980. Financial intermediation in economic development.

Dale F. Kuntz, Ph.D. Connecticut 1980. Agriculture and economic development: A comparative analysis of India, Brazil, and Taiwan.

Manuel Lasaga, Ph.D. Pennsylvania 1979. Econometric analysis of the impact of copper on Chile.

Robert B. Litterman, Ph.D. Minnesota 1980. Techniques of forecasting using vector autoregressions.

Anuradha Luther, Ph.D. Stanford 1980. The developing economy—Rural-urban migration, urban unemployment, and wages: A steady-state reconciliation.

Malcolm McPherson, Ph.D. Harvard 1980. Essays on the employment problems in Zambia.

Stanley Mahlala, Ph.D. New School 1980. The economic transformation of indigenous agriculture in Zimbabwe, 1900-65.

William Makinen, Ph.D. Michigan 1979. A social benefit-cost analysis of anti-measles vaccination in Yaounde, Cameroon.

Darius H. Mans, Ph.D. Massachusetts Institute of Technology 1980. The economics of infant feeding practices in developing countries.

Denise E. Markovich, Ph.D. Manitoba 1979. Incomes policy and the rate of wage increase: A comparison of Canada and the United States, 1956-76.

Talala B. Mbise, Ph.D. Indiana 1980. Comparative analysis of economic structures in an input-output framework: Experiments with East European data.

Philip Mirowski, Ph.D. Michigan 1979. The birth of the business cycle.

Abdelrazig E. Muhammed, Ph.D. Michigan State 1979. Labor requirements and supply for cotton picking with special reference to tenant family labor in New Halfa Agricultural Corporation, Sudan.

John Ndebbio, Ph.D. Oklahoma State 1980. Absorption capacity and growth rates in Nigeria, 1961-76.

Chiselebwe Ng'Andwe, Ph.D. Connecticut 1980. Inflation in Zambia: 1964-76.

Quoc-Trung Ngo, Ph.D. Wisconsin (Madison) 1980. An economic analysis of water resources development in deltaic regions of Asia: The case of central Thailand.

Joseph V. Ntangsi, Ph.D. California (Berkeley) 1979. The political economy of rural development in Cameroon.

Jeremiah Ojowu, Ph.D. Connecticut 1980. Technological transformation in Nigerian agriculture: An analysis of the Nigerian agricultural development policies, including the role of the marketing boards.

James O. Olukosi, Ph.D. Michigan State 1979. The distribution of personal incomes among African farmers: A two-period analysis.

Raghab D. Pant, Ph.D. Carleton 1980. A theoretical and empirical analysis of the sources of inflation: The Asian experience.

Joe B. Peek, Ph.D. Northwestern 1979. Inflation, capital gains, and saving behavior.

Mauricio B. Pinto, Ph.D. Johns Hopkins 1980. Brazilian manufacturing exports: Growth and change in structure.

Rafael Posada, Ph.D. Purdue 1980. The impact of the high-yielding varieties on the import requirements of the LDCs: The case of wheat.

Gerald D. Reiss, Ph.D. California (Berkeley) 1980. Development of Brazilian industrial enterprise: An historical perspective.

Farhad Rezazadeh, Ph.D. Iowa State 1979. Agricultural development in Iran: Evaluation of state planning and policies in relation to agriculture.

Helena J. Ribe, Ph.D. Yale 1979. Income of migrants relative to nonmigrants in Colombia: An economic analysis of the 1973 Census sample.

David L. Roberts, Ph.D. Rice 1980. Life cycle income and consumption in Colombia.

Ehud Satt, Ph.D. Stanford 1979. Problems in the kibbutz economy.

John G. Selody, Ph.D. Western Ontario 1979. A monetary model of Canadian price and output fluctuations.

Yun Wing Sung, Ph.D. Minnesota 1979. Factor proportions and comparative advantage in a trade-dependent economy: The case of Hong Kong.

Osamu Takenaka, Ph.D. Purdue 1979. Essays on the theory of endogenous technological progress.

Fayez A. Tarawneh, Ph.D. Southern California 1980. Income inequality and poverty in Mexico: Quantitative assessments and explanations.

Sonia M. Teixeira, Ph.D. Purdue 1979. Income, family composition, and social factors as variables in an expenditure model: The case of Brazil.

Somchai Thepthana, Ph.D Kentucky 1979. Government expenditures, taxes, and income distribution in Thailand.

Andras W. Uthoff, Ph.D. California (Berkeley) 1980. The distribution of income in Chile, 1960-74.

Theodore R. Valentine, Ph.D. Illinios 1979. Government wage policy, wage determination, and the development process: A case study of the Tanzanian economy, 1951-76.

Sweder Van Wijnbergen, Ph.D. Massachusetts Institute of Technology 1980. Credit policy, inflation, and growth in less-developed countries.

Mohamad T. Vaziri, Ph.D. Oklahoma 1979. The impact of the transfer of technology of multinational corporations on the growth of the private sector in developing economies with special reference to Iran.

Rogerio L. F. Werneck, Ph.D. Harvard 1980. Rapid growth, distributional equity, and the size of the public sector: Tradeoffs facing the Brazilian economic policy in the 1980's.

Shahid Yusuf, Ph.D. Harvard 1980. Rural social relations and agricultural development in Pakistan.

Elsayed A. Zaki, Ph.D. Michigan State 1980. An on-going evaluation of the planning, implementation, and tenancy (farm) size of the Rahad Irrigation Project of the Sudan.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

Badi H. Baltagi, Ph.D. Pennsylvania 1979. Econometric models for pooling time-series of cross sections.

John J. Beggs, Ph.D. Northwestern 1980. Pooling time-series and cross-sectional data in dynamic econometric models.

Douglas Caves, Ph.D. Wisconsin (Madison) 1980. Measuring productivity growth in the U.S. railroad industry with an estimate of losses resulting from economic regulation.

Kalyan Chatterjee, D.B.A. Harvard 1979. Interactive decision problems with differential information.

Marsha J. Courchane, Ph.D. Northwestern 1980. Comprehensive income and consumption.

Luis E. Derbez, Ph.D. Iowa State 1980. Econometric modeling and decision support systems: An applied exercise for a Mexican firm.

Terry E. Dielman, Ph.D. Michigan 1979. Analysis of pooled cross-sectional and time-series data with an application to finance.

Ian R. C. Eggleton, Ph.D. Chicago 1979. Patterns, prototypes, and predictions.

William F. Finan, Ph.D. Pennsylvania 1979. The national impact of an oil embargo: An application of econometric modeling.

Patricia Fulton, Ph.D. Western Ontario 1980. Homeowner imputed rent and capital gains: Implications for equity within the tax system.

Saad A. A. Hassanein, Ph.D. Catholic 1980. On the explanatory power of econometric models of investment behavior: Analysis and comparison.

John M. Howell, Ph.D. California (Santa Barbara) 1979. Information costs and innovation kurtosis.

Ian J. Irvine, Ph.D. Western Ontario 1979. An analysis of the distribution of lifetime purchasing power in Canada.

Zvi Itzkovich, Ph.D. Pennsylvania 1979. Estimation of a dynamic expenditure system using cross-section and time-series data.

Wayne H. Joerding, Ph.D. Northwestern 1979. Lifetime consumption, labor supply, and fertility: A complete demand systems approach.

Kyle T. Johnson, Ph.D. Stanford 1980. Specific cost-of-living measurement.

HEEJOON KANY, Ph.D. Washington 1980. Four essays in mathematical economics and econometrics.

KENNETH C. LAITINEN, Ph.D. Chicago 1979. A theory of the multiproduct firm.

KAY KEUN-HYUN LEE, Ph.D. Rutgers 1979. The effects of technological developments on international trade: The case of the U.S. and Japanese steel industries.

GEORGE MCCOLLISTER, Ph.D. California (San Diego) 1979. New methods for identifying time-series models and an evaluation of the economic indicators.

RAFI MELNICK, Ph.D. California (Berkeley) 1979. An econometric study of consumption with unobservables.

H. DENNIS OBERHELMAN, Ph.D. Purdue 1979. Methods of estimation in temporal cross-sectional studies.

WILLIAM R. PARKE, Ph.D. Yale 1979. An algorithm for full-information estimation.

JAMES M. PRICE, Ph.D. North Carolina (Chapel Hill) 1979. Some tests for comparing estimators under the generalized mean square error criterion.

LEE H. REYNOLDS, Ph.D. Nebraska (Lincoln) 1980. An interactive optimization/decision approach to corporate planning: A bank management application.

MARTIN J. RINGO, Ph.D. Brown 1980. An investigation in the finite sample properties of a test statistic associated with seemingly unrelated regression equations estimation.

CHARLES P. SCHMIDT III, Ph.D. Chicago 1979. Homotopies and the existence and the computation of solutions of systems of non-linear equations.

JAIME J. SERRA-PUCHE, Ph.D. Yale 1979. A computational general equilibrium model for the Mexican economy: An analysis of fiscal and commercial policies.

JOHN L. SILVER, Ph.D. Duke 1980. Frequency domain estimation of error components models.

MARK D. SOSKIN, Ph.D. Pennsylvnia State 1979. Detection and treatment of multicollinearity with application to problems of estimating pollution costs from housing market models.

THUAN V. TRUONG, Ph.D. Kentucky 1980. Estimation of regression parameters in linear regression model with autocorrelated errors.

JACOB W. ULVILA, D.B.A. Harvard 1979. Decisions with multiple objectives in integrative bargaining.

ARTHUR D. WARGS, Ph.D. Michigan 1980. Risk instability in the electric utility industry.

Monetary and Fiscal Theory, Policy, and Institutions

KAZUMI ASAKO, Ph.D. Yale 1979. An essay on the announcement impact and the role of stabilization policy.

STEVEN N. BRAUN, Ph.D. Pennsylvania 1980. Flexing the flexible accelerator.

JOHN BRODE, Ph.D. Harvard 1980. Three essays on the mathematics of constructing and forecasting time-ordered economic models: Option pricing models.

DAVID A. T. BURTON, Ph.D. Western Ontario 1979. Expectations and a small, open economy.

SANG-JINH BYUN, Ph.D. Southern Methodist 1979. Theory of a banking firm with stochastic elements.

ROBERT E. CHATFIELD, Ph.D. Purdue 1979. An economic analysis of the holder-in-due-course trade regulation rule.

SAM B. COOK, Ph.D. Missouri (Columbia) 1980. The performance and competitive impact of multibank holding companies in Missouri.

ROBERT H. DEFINA, Ph.D. Washington (St. Louis) 1979. Substitution of a federal value-added tax for the corporation income tax: An equilibrium analysis.

KEITH G. DUBLIN, Ph.D. Illinois 1979. A study of tax incidence in Guyana.

CONSTANCE R. DUNHAM, Ph.D. Stanford 1980. Financial innovation.

KEVIN G. DUNLEAVY, Ph.D. Duke 1979. The variability of the lagged effect of money on income.

GERALD DWYER, Ph.D. Chicago 1979. On the Gibson paradox.

ELYAS ELYASIANI, Ph.D. Michigan State 1979. Expected utility maximization and the theory of multiproduct banking firm under uncertainty.

JAMES M. FERRIS, Ph.D. North Carolina (Chapel Hill) 1979. The economic and demographic determinants of public goods demands: A systems approach.

ARLYN C. FREDRICK, Ph.D. Wisconsin (Milwaukee) 1979. Are government bonds net wealth?

LAWRENCE Y. FU, Ph.D. Rochester 1980. Money, inflation, and economic activity: A test of the dominant impulse hypotheses for the case of Taiwan.

GEARY A. GESSLER, Ph.D. California (Santa Barbara) 1979. Two essays on interest rates and inflation.

JAMES E. GLASSMAN, Ph.D. Northwestern 1979. The impact of Social Security on private saving.

AMNON GOLDSCHMIDT, Ph.D. Simon Fraser 1979. Liquidity production and the theory of the banking firm.

ABRAHAM Z. GOLDSTEIN, Ph.D. Yale 1979. Models of capital asset pricing with transaction costs.

NICOLAAS GROENEWOLD, Ph.D. Western Ontario 1979. The effectiveness of fiscal policy: A theoretical analysis.

JAMES J. GRUNLOH, Ph.D. Washington (St. Louis) 1979. The horizontal equity effects of equalizing aids in Wisconsin: An application of the median voter model.

WILLIAM S. HARAF, Ph.D. Washington 1979. The persistent effects of aggregate demand shocks.

OSKAR HARMON, Ph.D. Rutgers 1980. The incidence of the local property tax: A case study of New Jersey.

Yousef A. Hayajneh, Ph.D. Wayne State 1979. The demand for money in Japan: An econometric analysis.

Kathy J. Hayes, Ph.D. Syracuse 1980. The fiscal impact of tax abatement on commercial and industrial industries.

Charles E. Hegji, Ph.D. Washington (St. Louis) 1980. The intermediate target and information variable approaches to open market strategy.

Scott E. Hein, Ph.D. Purdue 1979. Empirical evidence on macro-economic money demand in the United States.

Zvi Hercowitz, Ph.D. Rochester 1980. Money and the dispersion of relative prices.

Yuzo Honda, Ph.D. Princeton 1980. A micro foundation for the transmission mechanism of monetary policies in Japan.

Seok-Hyun Hong, Ph.D. Stanford 1980. General equilibrium analysis of Korean taxation policy.

Frederic A. Hooper, Jr., D.B.A. Harvard 1979. Revenue variability and municipal debt.

Lawrence J. Horan, Ph.D. Columbia 1980. The forward market for loanable funds at life insurance companies.

Brian R. Horrigan, Ph.D. California (Los Angeles) 1980. An estimation and test of the Brunner-Meltzer model of inflation and output fluctuations.

Peter D. Ivory, Ph.D. California (Santa Barbara) 1979. Causality and the term structure of interest rates.

Douglas H. Joines, Ph.D. Chicago 1979. Government fiscal policy and private capital formation.

David S. Jones, Ph.D. Harvard 1980. A structural econometric model of the U.S. equity market.

Mark D. Karper, Ph.D. Cincinnati 1980. Predicting income velocity with money stock ratios: Theory and implication.

Joseph J. Kiernan, Ph.D. Fordham 1980. The determinants of the flow of savings to credit unions, 1965-76.

Dae Sik Kim, Ph.D. California (Santa Barbara) 1979. Some monetary implications of the government budget deficit in an open developing economy: A macro-economic simulation model of fiscal policy in Korea.

Kyung Moo Kim, Ph.D. Pennsylvania State 1979. The money supply process: Linkages between wage-price setting and central bank behavior.

Robert G. King, Ph.D. Brown 1980. Asset markets and the neutrality of money.

Susan F. Krause, Ph.D. Southern California 1980. Fully rational optimization, costly information, and the neutrality of money.

Shiro Kurihara, Ph.D. Johns Hopkins 1980. Imported inputs and macro-economic policy.

Patrick T. Laverty, Ph.D. Queen's 1979. Financial instruments and wealth.

Yang-Pal Lee, Ph.D. Stanford 1980. Inflation hedges and economic growth in a monetary economy.

Neill B. McGrath, Ph.D. Brown 1980. The government budget constraint and the effects of government financing.

Scott McKinney, Ph.D. Indiana 1979. Jurisdictional public service spillovers and interagency cooperation: The effects on police service expenditures and output.

Jorge L. Martinez, Ph.D. Washington (St. Louis) 1979. Benefit and tax incidence at the local public finance level.

Stephen A. Mathis, Ph.D. Iowa State 1979. An approach to the delineation of geographical banking market areas.

Kenneth T. Mayland, Ph.D. Pennsylvania 1979. A study of the consumer finance industry.

Mark W. Meador, Ph.D. Washington 1980. Determinants of interregional differentials in home mortgage rates.

Frederick C. Meltzer, Ph.D. Virginia 1979. Market organization and the management of investment funds.

Bahman Mirshab, Ph.D. Wayne State 1979. A case for proportional federal personal income tax.

Andrew J. Moody, Ph.D. Boston College 1979. Impact of state tax policy on manufacturing activity in Massachusetts.

Alan J. Norton, Ph.D. Missouri (Columbia) 1980. Money market certificates and the demand for savings and small time deposits.

Patrick T. Owens, Ph.D. Queen's 1980. A model of the term structure of interest rates.

Turgay O. Özkan, Ph.D. Massachusetts Institute of Technology 1980. Rational expectations and the role of monetary policy: A game-theoretic approach.

Anthony V. Popp, Northern Illinois 1980. The optimal provision of public goods in a representative democracy.

Javier Quesada, Ph.D. Cincinnati 1980. A model of the term structure of interest rates.

Perry D. Quick, Ph.D. Stanford 1979. Inflation and financial intermediaries: Effects on large banks, 1968-75.

Alberto Romez-Rivas, Ph.D. Rice 1980. Incidence of fiscal policies by input-output analysis.

Rusdu Saracoglu, Ph.D. Minnesota 1980. Two essays on interest rates and inflation.

Edward P. Shea, Ph.D. Boston College 1980. Large bank liability management behavior and the market for immediately available funds.

Joel B. Slemrod, Ph.D. Harvard 1980. A general equilibrium model of capital income taxation.

Charles E. Storrs, Jr., Ph.D. Georgia State 1979. The effect of banking market structure upon the transmission of monetary policy.

James H. Vanderhoff, Ph.D. North Carolina (Chapel Hill) 1979. An investigation of the formation of inflationary expectations.

Douglas G. Waldo, Ph.D. North Carolina (Chapel Hill) 1980. Long-term contracts and monetary policy.

Earl D. Wasson, Ph.D. Oklahoma State 1980. A micro-simulation analysis of changes in Oklahoma's individual income tax code.

James A. Wilcox, Ph.D. Northwestern 1980. Excess reserves in the interwar period.

Thomas O. Wisley, Ph.D. Purdue 1979. An experiment with mandatory balanced federal budgets.

Jisung Yoo, Ph.D. Catholic 1980. Money as an input factor: An econometric treatment of factor substitutability.

James Zoda, Ph.D. Northern Illinois 1979. The demand for money in the United Kingdom, 1948-76.

George R. Zodrow, Ph.D. Princeton 1980. Optimal tax reform: The case of property tax equalization.

International Economics

Teshome Abebe, Ph.D. Northern Illinois 1980. An econometric study of the determinants of international patterns of trade: OECD and Africa.

M. Shahid Alam, Ph.D. Western Ontario 1979. Welfare implications of trade policies: Three essays.

Tsehai Alemayehu, Ph.D. Kentucky 1979. A quantitative analysis of the factors determining the effectiveness of devaluation for balance-of-trade adjustment.

Michael D. Andrews, Ph.D. Wisconsin (Madison) 1979. The impact of market power and exchange rate varation on the profitability of devised multinational corporations.

George A. Apenteng, Ph.D. Pittsburg 1979. Government policy and the balance of payments: A case study of Ghana, 1965-74.

Nizamettin Aydin, Ph.D. Michigan 1980. The transfer of marketing know-how by multinational companies: A case study of Turkey.

Lakhindar Banerji, D.B.A. Harvard 1979. Empirical studies of exchange market behavior in the floating regime, 1973-78.

Dallas S. Batten, Ph.D. Ohio State 1980. An explanation of reciprocity and its relationship to bilateralism.

Charles Bende, Ph.D. Northern Illinois 1979. U.S. industrial metal imports: Origin, stability, and differential responses to price and exchange rate influences.

Benjamin Bental, Ph.D. Minnesota 1979. Capital controls in a two-sector model of growth and trade.

Jagdeep S. Bhandari, Ph.D. Southern Methodist 1980. The macro dynamics of international exchange rates.

Harry P. Bowen, Ph.D. California (Los Angeles) 1980. An empirical analysis of the product cycle theory of international comparative advantage: The global approach.

Michael D. Buckley, Ph.D. Wayne State 1979. Short-term capital movements in the U.S. balance of payments and commercial bank portfolio management.

Philippe Callier, Ph.D. Simon Fraser 1979. The covered interest arbitrage margin: An investigation into the relevance of the modern theory of the forward exchange rate.

David S. Clifton, Ph.D. Georgia State 1980. An empirical investigation of the Heckscher-Ohlin theorem.

J. Markham Collins, Ph.D. Oklahoma 1980. The forward market for foreign exchange: Characteristics and efficiency.

Gary Craig, Ph.D. Chicago 1979. A montary approach to the balance of trade.

Robert E. Cumby, Ph.D. Massachusetts Institute of Technology 1980. Multiple tier exchange rates and international capital mobility.

Jorge A. B. de Macedo, Ph.D. Yale 1979. Portfolio diversification and currency inconvertibility: Three essays in international monetary economics.

Shidan Derakhshani, Ph.D. Harvard 1980. Structuring international transfers of technology: Lessons from Iran.

Robert S. Dohner, Ph.D. Massachusetts Institute of Technology 1980. Essays on terms of trade changes and international adjustment.

Barry J. Eichengreen, Ph.D. Yale 1979. Tariffs and flexible exchange rates: The case of the British General Tariff of 1932.

Mark G. Farah, Ph.D. Oregon 1980. The competitiveness of U.S. exports of manufactures to Third World countries.

Victoria S. Farrell, Ph.D. Wisconsin (Madison) 1979. Optimal pricing policy for the OPEC cartel and the effects of non-OPEC combative tariff policy and U.S. price controls thereon.

Robert A. Feldman, Ph.D. Wisconsin (Madison) 1979. The impact of oil import price shocks: The interactive effects of domestic price controls, the exchange rate and macro policy on domestic prices.

Robert P. Ford, Ph.D. Queen's 1980. Three essays on exchange rate determination.

David J. Goldsbrough, Ph.D. Harvard 1979. Preferential trade arrangements amongst less-developed countries.

Marvin S. Goodfriend, Ph.D. Brown 1980. Money, the terms of trade, and balance-of-payments adjustment with a flexible exchange rate: A theoretical and empirical study of the dynamics of international adjustment.

Gene M. Grossman, Ph.D. Massachusetts Institute of Technology 1980. Essays on import competition and commercial trade policies.

Allan Gruchy, Ph.D. Wisconsin (Madison) 1980. Multisector dynamics for complete economics.

Brian H. Guck, Ph.D. Clark 1980. International transmission of inflation under fixed and floating exchange rate systems.

Sung Keun Ha, Ph.D. Indiana 1979. Monetary causal models for the balance of payments and exchange rates.

CRAIG HAKKIO, Ph.D. Chicago 1979. Expectations and the foreign exchange market.

JAMES C. HARTIGAN, Ph.D. Duke 1979. Trade policy and income distribution in the United States and Canada.

JAMES P. HEALY, Ph.D. Princeton 1980. The asset market determination of exchange rates in a multi-country setting.

DOUGLAS R. HEINSON, Ph.D. Iowa 1979. The international transmission of disturbances under flexible exchange rates.

RONALD R. HELGENS, Ph.D. Iowa 1979. A rational expectations model of the monetary approach to the balance of payments.

DALE W. HENDERSON, Ph.D. Yale 1979. Macroeconomic policy making in open economies: Four essays based on the portfolio balance approach.

MARK G. HERANDER, Ph.D. North Carolina (Chapel Hill) 1979. A theoretical and empirical analysis of the economic impact of government discrimination against foreign suppliers in procurement.

MARC W. HEROLD, Ph.D. California (Berkeley) 1979. An analysis of the historical determinants of the U.S. investment in Latin American manufacturing.

ABRAHAM J. HOLLANDER, Ph.D. Minnesota 1979. The foreign location decisions of U.S. transnational firms.

MUNAWAR IQBAL, Ph.D. Simon Fraser 1979. The effects of the EEC and the EFTA on trade flows: A cross-sectional analysis.

SHAFIQUL ISLAM, Ph.D. Harvard 1980. Essays on the balance of payments, devaluation, and development.

JAYANTI L. JAIN, Ph.D. Oklahoma State 1980. An empirical study on world money supply, gross world output, and inflation.

J. PETER JARRETT, Ph.D. Harvard 1979. Offshore assembly and production, and the internalization of international trade within the multinational corporation: Their causes and effects on U.S. manufacturing industry wage and profit rates.

RONALD A. JOHNSON, Ph.D. Stanford 1980. The scale of government in a small, open economy.

VIJAY K. JOLLY, Ph.D. Harvard 1979. Transformation regimes, product cycle distortions, and the export of manufactures from large developing countries.

C. MICHAEL JONES, Ph.D. Yale 1979. Policymaking efficiency and the international monetary system.

CHUNGSOO KIM, Ph.D. Johns Hopkins 1980. Changing pattern of comparative advantage: A neo-factor-proportions approach.

STUART C. KIRK, Ph.D. Claremont 1979. International market warfare: A contribution to the theory of economic diplomacy.

O. YOUNG KWON, Ph.D. New York 1980. Theory of foreign exchange and economic policy.

THOMAS A. LAYMAN, Ph.D. North Carolina (Chapel Hill) 1979. The economics of export controls—with an empirical application to the Brazilian coffee industry.

KAI-CHEONG LEE, Ph.D. Pennsylvania 1980. Macroeconometric studies on foreign trade and economic growth: The case of Taiwan and South Korea with discussions on further development, with the hindsight of Japan's experience.

CHI-HUNG LEUNG, Ph.D. Johns Hopkins 1980. The design of stabilization policies and the international transmission of disturbances.

JANE-YU HO LI, Ph.D. Wisconsin (Madison) 1980. Short-run demand for imports and domestic substitutes in the United States.

JOSE S. LIZONDO, Ph.D. Chicago 1979. Three essays in international economics.

DAVID L. LONGWORTH, Ph.D. Massachusetts Institute of Technology 1980. Floating exchange rates: The Canadian experience.

SUBODH C. MATHUR, Ph.D. Massachusetts Institute of Technology 1980. The economics of the commodities fund: An aspect of the new international economic order.

CHRISTOPHER L. MILLER, Ph.D. California (Berkeley) 1979. The role of the nation-state in foreign direct investment.

CRAIG MOORE, Ph.D. Syracuse 1980. Transfer pricing in multinational enterprises: Empirical tests.

MAURICE M. OBSTFELD, Ph.D. Massachusetts Institute of Technology 1980. Capital mobility and monetary policy under fixed and flexible exchange rates.

MARIANNE ODJAGOV, D.B.A. Harvard 1978. The foreign ownership policies of U.S. multinational banks.

THOMAS P. O'TOOLE, Ph.D. Johns Hopkins 1980. Inflation, flexible exchange rates, and macro-economic policy.

ABDESSATAR OUANES, Ph.D. Minnesota 1979. Econometric study of the international transmission of inflation: The case of West Germany.

AMAR K. PARAI, Ph.D. Southern Methodist 1979. Multinational corporations and international allocation of resources: A study in the theory of international trade and direct investment.

ARIS A. PROTOPAPADAKIS, Ph.D. Chicago 1979. Money, prices, and exchange rates in the German hyperinflation: An expectations approach.

WISARN PUPPHAVESA, Ph.D. Hawaii 1980. Export instability in ASEAN countries.

WILLIAM J. RIEBER, Ph.D. Pittsburgh 1979. Commodity market monopoly and customs unions.

KENNETH S. ROGOFF, Ph.D. Massachusetts Institute of Technology 1980. Essays on expectations and exchange rate volatility.

JEFFREY D. SACHS, Ph.D. Harvard 1980. Factor costs and macro-economic adjustment in the open economy: Theory and evidence.

ALVARO-JOSE SAIEH-BENDECK, Ph.D. Chicago 1980. Behavior of monetary variables in fixed exchange rate economies.

MARSHA R. SHELBURN, Ph.D. North Carolina (Chapel Hill) 1979. A comparison of alternative regulations designed to limit foreign exchange market intervention by national authorities during a managed float.

Chang Min Shin, Ph.D. Claremont 1980. Capital imports in economic development: The Korean case, 1962-76.

Edward F. Stuart, Ph.D. Oklahoma 1980. Wage differentials and foreign trade.

Rene M. Stulz, Ph.D. Massachusetts Institute of Technology 1980. Essays on international asset pricing.

Toshiko Tange, Ph.D. Pennsylvania 1979. Changes in technology, costs, and in comparative advantage: An empirical study of Japanese-U.S. manufacturing exports.

Serif C. Ulusoy, Ph.D. Pittsburgh 1980. Reaction functions and sterilization coefficients.

Massoumeh S. Vafair, Ph.D. Cincinnati 1979. Welfare effects of Iranian trade liberalization.

Jean-Marie Viaene, Ph.D. Pennsylvania 1980. Customs union between Spain and EEC: An attempt at quantification of the long-term effects in a general equilibrium framework.

Keith Worrell, Ph.D. Simon Fraser 1979. How money matters in an open developing economy: A case study of Jamaica, 1962-71.

Mickey T. C. Wu, Ph.D. Kentucky 1979. An empirical analysis of a monetary approach to the balance of payments and exchange rates: The case of Singapore.

Miranda Xafa, Ph.D. Pennsylvania 1979. EEC and Greece: An econometric analysis of the effects of association on imports of manufactures, 1962-72.

Joon Hak Yoo, Ph.D. Minnesota 1979. An estimate of the welfare changes from formation of a customs union among some countries in Southeast Asia.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting.

Stephen L. Able, Ph.D. Indiana 1979. Aggregate investment behavior, the neoclassical theory, and uncertainty: An empirical study.

Rowland K. Atiase, Ph.D. California (Berkeley) 1980. Predisclosure, informational asymmetries, firm capitalization, financial reports, and security price behavior.

Roger J. Baran, Ph.D. Chicago 1980. Family financial decision making.

Christopher A. Bartlett, D.B.A. Harvard 1979. Multinational structural evolution: The changing decision environment in international divisions.

John E. G. Bateson, D.B.A. Harvard 1980. The problems of research utilization in marketing: A pilot study of the research supplier selection process in consumer goods companies.

Richard A. Bettis, Ph.D. Michigan 1979. Strategic management in the multibusiness firm: Implementing the portfolio concept.

Gary C. Biddle, Ph.D. Chicago 1980. Accounting methods and management decisions: The case of inventory costing and inventory policy.

Paul D. Borge, D.B.A. Harvard 1980. Long-run consumption and investment policies: Adaptive strategies versus uniform rules.

Michael Bradley, Ph.D. Chicago 1979. An analysis of interfirm cash tender offers.

John A. Cole, Ph.D. Michigan 1980. Theory and evidence regarding the information content in dividends.

William H. Davidson, D.B.A. Harvard 1979. Corporate experience factors in international investment and licensing activity: A study of the international spread of manufacturing for products developed in the United States.

Douglas V. DeJong, Ph.D. Michigan 1980. The value of the firm and its obligations: Implications for accounting.

Douglas Diamond, Ph.D. Yale 1980. Essays on information and financial intermediation.

Richard R. Ellsworth, D.B.A. Harvard 1980. Corporate strategy and capital structure policies: A descriptive study.

Robert Evans, Ph.D. Wisconsin (Madison) 1980. On the existence, measurement, and economic significance of market pressure in the pricing of new equity shares.

Cynthia J. Frey, Ph.D. Michigan 1980. An information integration theory approach to the study of consumers' cognitive algebra.

David A. Gautschi, Ph.D. California (Berkeley) 1979. Models of individual retail center choice.

John P. Goldsberry III, Ph.D. Harvard 1980. Dynamic models of corporate debt management.

Anil K. Gupta, D.B.A. Harvard 1980. The process of strategy formation: A descriptive analysis.

Judith B. Kamm, D.B.A. Harvard 1980. The balance of innovative behavior and control in new product development.

Charles W. Kennedy, Ph.D. Michigan 1980. An empirical comparison of several within-subjects terms of expectancy theory.

Thomas E. Kilcollin, Ph.D. Chicago 1980. Retail advertising and media mix.

Duncan G. LaBay, Ph.D. Michigan 1980. An exploration of the consumer decision process in the adoption of solar energy systems.

Leslie Levy, D.B.A. Harvard 1980. Board leadership and the chairman's role in the large publicly held corporation.

Anthony K. Lima, Ph.D. Stanford 1980. Inflation and the demand for life insurance.

Jeanne M. Lynch, D.B.A. Harvard 1979. Activating the board of directors: A study of the process on increasing board effectiveness.

Rowland T. Moriarty, D.B.A. Harvard 1980. The use of organizational buying in assessing industrial markets.

Philip Nelson, Ph.D. Yale 1980. Bankruptcy: A behavioral theory of the firm approach.

Philip R. Perry, Ph.D. California (Berkeley) 1979. Properties of extreme security returns.

Allan Popoff, Ph.D. Queen's 1980. A dynamic econometric model of corporate financing behavior.

V. Jean Ramsey, Ph.D. Michigan 1979. Organizational structure of academic departments as a function of environmental uncertainty or task routineness: Methodological and measurement issues.

Marc R. Reinganum, Ph.D. Chicago 1979. Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values.

Daleen D. Richmond, Ph.D. Cornell 1980. An export commodity marketing computer simulation game.

Alfred Sant, D.B.A. Harvard 1979. Problems of government organization and management in the state enterprises of developing countries: The Maltese case.

Brian Scott, D.B.A. Harvard 1979. The organizational network: A strategy perspective for development.

Jonathan A. Scott, Ph.D. Purdue 1980. An empirical analysis of daily cash management optimization models.

P. David Shields, Ph.D. Michigan 1980. The CPA firm referral behavior of attorneys and bankers for a typical small business.

Dan A. Simunic, Ph.D. Chicago 1979. The pricing of financial audit services: Theory and evidence.

Brigitta Swedenborg, Ph.D. California (Los Angeles) 1979. The multinational operations of Swedish firms: An analysis of determinants and effects.

David Swenson, Ph.D. Yale 1980. A model for the valuation of corporate bonds.

John R. Twombly, Ph.D. Chicago 1980. The information content of segment reports.

Srinivasan Umapathy, D.B.A. Harvard 1980. The process of budgeting in decentralized firms.

Ronald Utt, Ph.D. Indiana 1980. The influence of transactions costs on the existence of financial intermediaries.

Terrence H. Witkowski, Ph.D. California (Berkeley) 1980. Determinants of advertising regulation attitudes: The alcoholism field and alcohol advertising.

Robert F. Young, D.B.A. Harvard 1980. The uses and effectiveness of vertical cooperative advertising.

Daniel M. Zeghal, Ph.D. Laval 1979. The effect of firm size on the information value of financial statements.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

Zoltan Acs, Ph.D. New School 1980. Price behavior and the theory of the firm in competitive and corporate markets.

Richard G. Anderson, Ph.D. Massachusetts Institute of Technology 1980. Energy conservation and factor substitution in U.S. manufacturing, 1947-71.

David L. Baumer, Ph.D. Virginia 1980. Federal regulation of the dairy industry: Costs, benefits, and legal constraints.

Joe A. Bell, Ph.D. Oklahoma State 1980. Cost of regulation in commercial banking: The tenth Federal Reserve district.

Paul Biederman, Ph.D. New School 1980. The domestic airline industry: The end of an era.

Jerome T. Bock, Ph.D. California (Berkeley) 1980. The value of reliability and optimal price/capacity decisions for public and private utilities.

Dan C. Boger, Ph.D. California (Berkeley) 1979. Economic aspects of the regulated inland waterway freight transportation industry.

Timothy F. Bresnahan, Ph.D. Princeton 1980. Three essays on the American automobile oligopoly.

Jonathan L. S. Byrnes, D.B.A. Harvard 1980. Air carrier diversification.

Salim U. Chishti, Ph.D. Washington (St. Louis) 1979. Productivity and capacity utilization in the manufacturing sector of Pakistan.

William A. Coleman, Ph.D. Duke University 1979. A structural model of symphony orchestra firms: Theory, estimation, forecasts, and policy evaluations.

Ronald B. Conners, Ph.D. New School 1980. The growth of hospital care.

Timothy W. Cooke, Ph.D. Johns Hopkins 1980. Intraurban industrial location: A theoretical and empirical analysis.

Janna L. Cowen, Ph.D. Nebraska (Lincoln) 1979. The operating ratio and alternative earnings control standards for regulated motor freight carriers.

Carl J. Dahlman, Ph.D. Yale 1979. A micro-economic approach to technical change: The evolution of the Usiminas steel firm in Brazil.

Robert Damus, Ph.D. Chicago 1979. A two-part Ramsey-optimum railroad tariff.

James Dechene, Ph.D. Michigan 1979. Multiparty accidents: Choices, incentives, and optimal liability rules.

Daniel-Guy Denoual, D.B.A. Harvard 1980. The diffusion of innovations: An institutional approach.

Dionissios A. Dimopoulos, Ph.D. Columbia 1979. Pricing schemes for regulated enterprises and their implications in the case of electricity.

Patricia M. Dinneen, Ph.D. Massachusetts Institute of Technology 1980. Departures from efficiency in regulated industries: The case of international telecommunications.

Jay P. Ford, Ph.D. Washington 1979. Substitution among capital, labor, and energy inputs in U.S. manufacturing.

Merritt Fox, Ph.D. Yale 1980. The role of finance in industrial organization: A general theory and the case of the semiconductor industry.

LAYTON W. FRANKO, Ph.D. New School 1980. The economics of the network television industry.

JAMES R. FREW, Ph.D. Purdue 1979. Utilization of ICC operating rights prices to evaluate the impact of ICC regulation.

DOV FRISHBERG, Ph.D. Columbia 1980. Peak load pricing: Critical reviews, extension, and an application.

K. CELESTE GASPARI, Ph.D. Wisconsin (Madison) 1979. Exporting, foreign investment, and domestic market power.

HENRY GROSS, Ph.D. Columbia 1980. An economic evaluation of the legal doctrine of breach by anticipatory repudiation.

JOHN HANSEN, Ph.D. Yale 1980. Competitive aspects of the U.S. petroleum pipeline industry: Implications for regulatory policy.

BARRY C. HARRIS, Ph.D. Pennsylvania 1979. The multidivision form organization: Studies of its effects on performance in large organizations.

SCOTT M. HARVEY, Ph.D. California (Berkeley) 1979. Firm expansion behavior and the oil refining capacity "crisis" of 1973.

ERIC I. HEMEL, Ph.D. Stanford 1980. Rate base accounting and allocative efficiency for a regulated firm.

JOHN HOVEN, Ph.D. Wisconsin (Madison) 1980. Effects of occupational licensing rules on the number and incomes of licensed professionals.

GARY L. JACKSON, Ph.D. Massachusetts 1979. An empirical analysis of competition among independent and major brand retail gasoline outlets during the pre- and postembargo years, 1971-76.

ROBERT W. JOHNSON, D.B.A. Harvard 1980. The passage of the Investment Incentive Act of 1978: A case study of business influencing public policy.

ROGER JOHNSON, Ph.D. Northern Illinois 1980. Optimum seasonal pricing of North Atlantic air passenger transportation.

STEPHEN A. JOHNSTON, Ph.D. North Carolina State 1979. Construction work in progress in the rate base versus allowance for funds used during construction: An investigation of the effects on capital intensity in electric utilities.

JOSEPH P. KALT, Ph.D. California (Los Angeles) 1980. Federal regulation of petroleum prices: A case study in the theory of regulation.

GORDON V. KARELS, Ph.D. Purdue 1979. Firm behavior under uncertainty: The case of OCS petroleum lease sales.

STEPHEN KARLSON, Ph.D. Wisconsin (Madison) 1980. Industrial location and economic performance: Spatial competition involving a raw material and an econometric model of production in space.

CHRISTINA M. KELTON, Ph.D. Wisconsin (Madison) 1979. Postwar relocation dynamics of manufacturing industries in the United States.

ROY W. KENNEY, Ph.D. California (Los Angeles) 1979. The economics of block booking.

CHIKARA KOMURA, Ph.D. Pennsylvania State 1979. The supply of single family houses by the speculative builder under sales period uncertainity.

MARK E. MCBRIDE, Ph.D. Washington (St Louis) 1979. Federal Trade Commission policy toward vertical mergers: The case of cement and concrete.

ADRIENNE M. MCELWAIN, Ph.D. Duke 1980. The modelling of the time-of-day demand for electricity: An investigation of the Carolina Power and Light Residential Time-of-Day Pricing Experiment.

DONALD J. MARSHALL, Ph.D. California (Berkeley) 1980. A stochastic model of service contracts with application to an oil tankship market.

IVOR P. MORGAN, D.B.A. Harvard 1980. The impact of the Jones Act on selected U.S. industries.

STEVEN A. MORRISON, Ph.D. California (Berkeley) 1979. Optimal pricing and investment policies for airport landing areas.

HUGH M. NEARY, Ph.D. California (Berkeley) 1979. Labor-managed firms: Monopolistic competition, uncertainty, and implicit contracts.

JEFFRY M. NETTER, Ph.D. Ohio State 1980. Advertising as an anticompetitive device.

LOWELL H. ORREN, Ph.D. Minnesota 1979. A general equilibrium model of optimal rate-of-return regulation with an application to electricity.

STARR PAGE-MCMULLEN, Ph.D. California (Berkeley) 1979. Profits and the cost of capital: The U.S. domestic trunk airline industry, 1954-74.

BRIAN PECKHAM, Ph.D. Wisconsin (Madison) 1979. Economics and invention: A technological history of the U.S. corn refining industry.

OWEN R. PHILLIPS, Ph.D. Stanford 1979. Product bundling and optimal selling strategies for a two-product firm.

WILLIAM A. PHILLIPS, Ph.D. Florida State 1979. The performance of joint bidding ventures in competition for outer continental shelf petroleum.

RUSSELL W. PITTMAN, Ph.D. Wisconsin (Madison) 1979. The costs of water pollution control to the Wisconsin paper industry: The estimation of a production frontier.

ROGER L. PUPP, Ph.D. Washington (St. Louis) 1979. Cross subsidies in a medical care prepaid group practice setting.

WILLARD W. RADELL, JR., Ph.D. Illinois 1979. Scale effects of seasonal production in the U.S. raw cane sugar industry.

LYNNE SCHNEIDER, Ph.D. California (Los Angeles) 1980. The economic effects of the ban on broadcast cigarette advertising.

ZENA K. A. SELDON, Ph.D. Manitoba 1979. The economic implications of alternative air transport regulatory practices: A Canada-United States comparison.

ISRAEL SHAKED, D.B.A. Harvard 1980. A dynamic analysis of financial decisions and growth under alternative patterns of entry threat.

Eui Soon Shin, Ph.D. Washington (St. Louis) 1980. Impact of the 1973-74 oil crisis and energy substitution in Korea.

Richard Silkman, Ph.D. Yale 1980. The diffusion of educational innovations.

Kriss Sjoblom, Ph.D. Yale 1980. Essays on information and regulation by price and on induced innovation and the rate of return to investment.

David D. Smith, Ph.D. Michigan 1979. The relationship between market structure and the effective corporate income tax rate.

Janet K. Smith, Ph.D. California (Los Angeles) 1980. Occupational and business licensing legislation: An economic analysis.

Lee I. Sparling, Ph.D. California Institute of Technology 1980. Regulatory distortions in transportation and telecommunications.

Pablo Spiller, Ph.D. Chicago 1980. Quality, capacity, and regulations of the airline industry.

Robert S. Stillman, Ph.D. California (Los Angeles) 1980. Examining antitrust policy towards horizontal mergers.

Ephraim Toaff, Ph.D. Carnegie-Mellon 1980. Models of product diversity in large markets with imperfectly informed buyers.

Octavio A. F. Tourinho, Ph.D. California (Berkeley) 1979. The valuation of reserves of natural resources: An option pricing approach.

Barry Vann, Ph.D. Michigan 1979. Risk aversion in the bidding for oil and gas leases on the outer continental shelf.

Thomas G. Watkins, Ph.D. Iowa State 1979. An analysis of sealed bidding practices in a public sector market.

Alix E. Werth, Ph.D. Massachusetts Institute of Technology 1980. The effects of regulatory policy on the cost of equity capital and the value of equity in the electric utility industry.

Alice P. White, Ph.D. Yale 1980. The creation, maintenance, and erosion of market power: An analysis of the dominant firm.

Clifford Winston, Ph.D. California (Berkeley) 1979. A disaggregated qualitative mode-choice model for intercity freight transportation.

William C. Wood, Ph.D. Virginia 1980. Nuclear liability, nuclear safety, and economic efficiency.

Agriculture and Natural Resources

Ahmed H. Ahmed, Ph.D. Ohio State 1980. Recent performance of rural financial markets in the Sudan, 1965-79.

Valentine J. Akenda-Ondoga, Ph.D. Cornell 1980. An economic study of production organization and labor use among smallholders of Nile Province, Uganda.

Hossein Akhavipour, Ph.D. Kansas State 1979. Assessment of benefits of Kansas water quality management plan.

Joyce E. Allen, Ph.D. Illinois 1980. An analysis of the Food Stamp Program: Policy decisions and household food expenditure patterns.

John Antle, Ph.D. Chicago 1980. Human capital infrastructure and technology choice in agricultural development.

Jeffrey D. Apland, Ph.D. Purdue 1979. A production-based analysis of regional crop residue supply for energy generation.

David Appel, Ph.D. Rutgers 1980. Estimating the benefits of air quality improvement: A hedonic price index approach applied to the New York metropolitan area.

Tulio Arvalo-Duran, Ph.D. Chicago 1979. Brazilian government policies in agriculture: The case of grains and soybeans.

Edinaldo G. Bastos, Ph.D. Cornell 1980. Farming in the Brazilian Sertao: Social organization and economic behavior.

Shashanka Bhide, Ph.D. Iowa State 1980. A separable programming model incorporating linear demand functions for grains and vegetable oils: An analysis of U.S. agriculture in 1985.

Jose V. Biserra, Ph.D. Ohio State 1980. Uncertainty and decision analysis on large sharecropped farms in northeast Brazil.

William G. Boggess, Ph.D. Iowa 1979. Development and application of an interregional separable programming model of U.S. agriculture in 1985.

John B. Braden, Ph.D. Wisconsin (Madison) 1979. Natural gas from the Great Lakes: Economics, environment, and policies.

Harry E. Brautigam, Ph.D. Illinois 1980. Analysis of objectives of the development of the agricultural frontier of Nicaragua.

Leslie J. Butler, Ph.D. Michigan State 1979. Some aspects of agricultural price instability.

Wu-Hsiung Chen, Ph.D. Illinois 1980. An economic study on government rice stock operation in Taiwan.

Gurprit Chhatwal, Ph.D. Kansas State 1980. Disincentive effect P.L. 480 wheat imports for India.

David L. Chicoine, Ph.D. Illinois 1979. Farmland values in an urban fringe: An analysis of market data from Will County, Illinois.

Ashok K. Chowdhury, Ph.D. Iowa State 1980. Competitive partial equilibrium analysis of U.S. agriculture under alternative energy and export situations: Application of separable and chance constrained programming.

Eric W. Crawford, Ph.D. Cornell 1980. A programming-simulation study of constraints affecting the long-run income-earning ability of traditional dryland farming systems in northern Nigeria.

Ronald E. Deiter, Ph.D. Illinois 1979. U.S. dairy cooperative service behavior: An overview of the

North Central region and a case study of the Chicago market.

Gorantla Doraswamy, Ph.D. Cornell 1979. Mechanization and its effect on the demand for and supply of agricultural labor in Chittoor District of India.

Edgardo D. Dosayla, Ph.D. Cornell 1979. A cross-section analysis of food consumption in the Philippines.

Daniel J. Dudek, Ph.D. California (Davis) 1979. Dynamic analysis and control of agricultural land use.

Seung Young Earmme, Ph.D. Alberta 1979. A water use projection model for the North Saskatchewan River Basin, Alberta, 1980-85: An input-output approach.

Gary P. Ender, Ph.D. Cornell 1980. The development of road transportation in Nepal and its relationship to agricultural development.

Francis M. Epplin, Ph.D. Iowa State 1979. Estimation and economic analysis of protein and energy utilization by beef steers.

Merlin W. Erickson, Ph.D. Nebraska (Lincoln) 1979. The economic impact of energy prices on crop production in southwest Nebraska.

Marvin Feldman, Ph.D. Wisconsin (Madison) 1979. Applications of decision analysis to Wisconsin 1979 state energy conservation plan.

Marco A. Ferroni, Ph.D. Cornell 1980. The urban bias of Peruvian food policy: Consequences and alternatives.

Andrew W. Foshee, Ph.D. Louisiana State 1980. Agrarian political economy: The order of ends and means.

Stephen C. Gabriel, Ph.D. Illinois 1979. Financial responses to changes in business risk on central Illinois grain farms.

Peter N. Gardner, Ph.D. British Columbia 1980. Optimal management of the Fraser River sockeye salmon.

Colin Gellatly, Ph.D. North Carolina State 1979. Incorporating futures prices in commodity models: The U.S. broiler sector.

William Grisley, Ph.D. Illinois 1980. Effect of risk and risk aversion on farm decision making: Farmers in northern Thailand.

Arthur H. Grube, Ph.D. North Carolina State 1979. Economic analysis of field scouting and insecticide use in cotton production.

Sulafa J. Hadeed, Ph.D. California (Santa Barbara) 1979. A study of new contractual alternatives for the disposal of oil resources in the Middle East oil-producing countries.

Bruce F. Hall, Ph.D. California (Berkeley) 1978. An economic analysis of multiple job holding by farm operators.

Michael C. Halse, D.B.A. Harvard 1979. The concept of conservation efficiency as applied to the Indian milk and food economy.

John D. Hanson, Ph.D. Iowa State 1980. An information theory analysis of grades and grading with an application to beef cattle.

Peter A. Hartmann, Ph.D. Illinois 1979. The forward pricing efficiency of the live-hog futures market.

Linda L. Hudgins, Ph.D. Hawaii 1980. An economic model of a fisheries market with endogenous supply: The Hawaii skipjack tuna case.

Patricia Hutton, Ph.D. Wisconsin (Madison) 1980. U.S. dairy policy: A quantitative analysis of the post-World War II period.

Donald A. Johnson, Ph.D. Iowa State 1979. Investment, production, and marketing strategies for an Iowa farmer-cattle feeder in a risky environment.

George M. Johnston, Ph.D. Michigan State 1980. Rural to urban land conservation.

Kandice H. Kahl, Ph.D. North Carolina State 1979. An analysis of intertemporal basis movements.

Amarjit Kahlon, Ph.D. Kansas State 1980. Regional interstate varations in supply response for rice in India.

Eddie H. Kaiser, Ph.D. Iowa State 1979. An economic evaluation of alternative financial strategies used by beginning farmers to enter agriculture.

Robert P. King, Ph.D. Michigan State 1979. Operational techniques for applied decision analysis under uncertainty.

Marvin L. Klein, Ph.D. Michigan State 1979. An information system approach to the study of price data: A case study of the Michigan potato industry.

Richard M. Klemme, Ph.D. Purdue 1980. An economic analysis of the on-farm grain handling decision problem.

Danny A. Klinefelter, Ph.D. Illinois 1979. Optimal marketing strategies in response to uncertainty by Illinois grain farmers.

Steven E. Kraft, Ph.D. Cornell 1980. Performance classification of New York soils: The farmer's perspective.

Seon Lee, Ph.D. Cornell 1980. Stabilization of world primary product markets: An application of optimal control theory.

Alfred Levinson, Ph.D. California (Berkeley) 1978. Energy and materials in three sectors of the economy: A dynamic model with technological change as an endogenous variable.

Soe Lin, Ph.D. Carleton 1980. A macro-policy simulation model of the Canadian agricultural sector.

Kerry K. Litzenberg, Ph.D. Purdue 1979. An analysis of aggregation bias in quadratic spatial equilibrium models: An application to federal milk marketing orders.

Richard C. McGinity, D.B.A. Harvard 1980. Technological change and agribusiness structure: The beef system.

J. Thomas McGuckin, Ph.D. Wisconsin (Madison) 1980. An economic assessment of alfalfa dewatering technology in Wisconsin.

Juvir L. Mattuella, Ph.D. Ohio State 1980. Economic impact of alcohol production on agriculture in southern Brazil.

Judas T. G. Mendes, Ph.D. Ohio State 1980. Soybean marketing strategies: A portfolio analysis.

Alphonse L. Meyer, Ph.D. Purdue 1980. Obstacles to carcass grade and weight marketing of livestock.

Ndanusa B. Mijandadi, Ph.D. Cornell 1980. Production efficiency on farms in northern Nigeria.

Mark D. Newman, Ph.D. Michigan State 1980. An evaluation of the economic potential for coordination of export marketing by U.S. farmer cooperatives.

Kent D. Olson, Ph.D. Iowa State 1979. The resource structure of U.S. agriculture: An economic analysis.

Mahboobeh Ordoobadi, Ph.D. Purdue 1980. A linear programming model for analyzing energy-economy interactions at the national and global levels.

Patrick D. O'Rourke, Purdue 1979. Pricing policies, and the firm and spatial market.

Douglas H. Pachico, Ph.D. Cornell 1980. Small farmer decision making: An economic analysis of three farming systems in the hills of Nepal.

Donald E. Parker, Ph.D. Wisconsin (Madison) 1979. Irrigation project performance and water distribution in the less-developed countries.

Glenn D. Pederson, Ph.D. Michigan State 1980. A conceptualization and analysis of the distributional impacts of alternative agricultural credit policies.

Philip W. Pepper, Ph.D. Illinois 1980. A U.S. corn stocks and consumption demand model.

Barry Primm, Ph.D. Kansas State 1980. An L.P. simulation analysis of agricultural and rural development for the northwestern Kansas counties of Decatur, Rawlins, and Sheridan.

Viswanatha Ramachandran, Ph.D. North Carolina State 1979. The economics of farm tractorization in India.

George W. Reeves, Ph.D. California (Davis) 1979. Price stabilization, bilateral trade, and institutional constraints: The case of beef in Australia and the United States.

Jeong J. Rhee, Ph.D. Iowa State 1980. The combination of input-output analysis and linear programming for water resource management: An application to northwest Iowa.

Roland K. Roberts, Ph.D. Iowa State 1979. Analysis of selected policy impacts on the U.S. livestock sector by a five-commodity econometric stimulation model.

James D. Schaub, Ph.D. North Carolina State 1979. A simultaneous equations model of multiple job-holding farmers with endogenous farm inputs.

Robert B. Schwart, Jr., Ph.D. Ohio State 1979. An economic analysis of fluid milk resale trade and price regulations in the thirteen North Central states.

Haim Shalit, Ph.D. California (Berkeley) 1978. Analysis of increasing farmland values.

John Siebert, Ph.D. California (Berkeley) 1978. Almonds, bees, and externalities in the California agricultural economy.

Gurnev Singh, Ph.D. Illinois 1979. Operating capital use in relation to investment in depreciable assets on farm firms in central Punjab, India.

Kenneth R. Stollery, Ph.D. Queens 1980. A theory of the exploration and production of minerals.

Loren W. Tauer, Ph.D. Iowa State 1979. A liability management model for federal intermediate credit banks.

David Terkla, Ph.D. California (Berkeley) 1979. The revenue capacity of effluent charges on air and water pollution.

Peter K. Thor, Ph.D. California (Davis) 1980. An economic analysis of the marketing orders for the California-Arizona orange industry.

Thomas F. Tice, Ph.D. Purdue 1979. An investigation of nitrogen application under uncertainty.

Eilif Trondsen, Ph.D. California (Santa Barbara) 1979. Discretionary resource allocation: An economic analysis of the Norwegian North Sea leasing system for oil and gas.

Jay D. Tucker, Ph.D. Michigan State 1979. The economic impact of rural road development on transportation assembly costs in agriculture.

Nelson J. Updaw, Ph.D. North Carolina State 1979. Market analysis of the component pricing of soybeans.

Henry N. Wallace III, Ph.D. California (Davis) 1979. Beet sugar in the United States: An econometric analysis of supply response.

Michael Wattleworth, Ph.D. California (Berkeley) 1980. Agricultural development through technological change and diversification: An economic analysis of the rubber estate sector of peninsular Malaysia, 1950's to 1970's.

Torng-Chuang Wu, Ph.D. California (Davis) 1979. The demand for food in Taiwan.

David Zilberman, Ph.D. California (Berkeley) 1979. A putty clay approach to environmental quality control.

Manpower, Labor, and Population; including Trade Unions and Collective Bargaining

Dennis A. Ahlburg, Ph.D. Pennsylvania 1979. Alternate approaches to forecasting U.S. fertility.

William T. Alpert, Ph.D. Columbia 1979. The economic determinants of private wage supplements.

Teresa L. Amott, Ph.D. Boston College 1980. Three essays on occupational segregation.

Jean M. Baldwin, Ph.D. Massachusetts Institute of Technology 1980. The response of other wages to the minimum wage: A theory and empirical work.

Richard Barichello, Ph.D. Chicago 1979. Rural farm schooling decisions: Canada.

Evan W. Barrington, Ph.D. Massachusetts Institute of Technology 1980. Mandatory retirement and the labor supply of older workers.

Mark E. Bayless, Ph.D. Washington (St. Louis) 1979. The differential attractiveness of locations and the spatial pattern of wages: The case of university professors.

Carl C. Brown, Ph.D. Oklahoma State 1979. An economic analysis of the GI Bill educational benefits: A study of Korean and post-Korean veterans.

James N. Brown, Ph.D. Chicago 1980. Employee risk aversion, income uncertainty, and optimal labor contracts.

Marian V. Brown, Ph.D. Stanford 1979. Retirement income and labor supply.

Richard J. Butler, Ph.D. Chicago 1979. Black/white wage growth: A look at South Carolina production workers.

John S. Cameron, Ph.D. Nebraska (Lincoln) 1979. U.S. professional manpower flows, financing, and inequalities.

Thomas J. Campbell, Ph.D. Chicago 1980. Earnings difference between men and women in federal government employment.

Harlod M. Cline, Ph.D. Princeton 1980. The impact of labor mobility and unemployment on the welfare of workers.

Donald C. Cox, Ph.D. Brown 1980. The effects of career interruptions on the earnings of women: A longitudinal study.

David L. Crawford, Ph.D. Wisconsin (Madison) 1979. Extimating models of earnings from truncated samples.

Gregory E. De Freitas, Ph.D. Columbia 1979. The earnings of immigrants in the American labor market.

James N. Dertouzos, Ph.D. Stanford 1979. Union objectives, wage determination, and the International Typographical Union.

Thomas M. Dickey, Ph.D. Michigan State 1980. The Puerto Rico wage rate subsidy for agricultural labor.

Juan Díez-Canedo-Ruiz, Ph.D. Massachusetts Institute of Technology 1980. A new view of Mexican migration to the United States.

Clifford B. Donn, Ph.D. Massachusetts Institute of Technology 1980. The history and economic policy of the Australian Council of Trade Unions.

Franklin T. Elder, Ph.D. Chicago 1980. Employment and labor force effects of changes in the federal minimum wage, 1966-78: A Box-Jenkins Approach.

Evangelos M. Falaris, Ph.D. Minnesota 1979. Migration: A study of choice among discrete alternatives.

Rudy H. Fichtenbaum, Ph.D. Missouri (Columbia) 1980. Earnings function specification and CETA program evaluation.

Errol A. Forde, Ph.D. Illinois 1979. Labor absorption in the service sector of a developing economy: The case of Puerto Rico.

Maria E. Freire, Ph.D. California (Berkeley) 1979. Assessing the economic contribution of education: A study for the Portuguese industrial sector.

George Fulton, Ph.D. Michigan 1979. A regional econometric analysis of the Detroit labor market area.

Otis W. Gilley, Ph.D. Purdue 1979. A theoretical and empirical investigation of the determinants of job search behavior.

James H. Grant, Ph.D. Michigan State 1979. Substitution among labor, labor and capital in U.S. manufacturing.

John C. Ham, Ph.D. Princeton 1980. Three empirical essays on constraints in the labor market.

Margaret L. Hashimoto, Ph.D. Columbia 1979. The effect of industrial composition changes on the relative demand for skilled labor.

Melba A. Hill, Ph.D. Duke 1980. Labor force participation of married women in urban Japan.

Martin Horning, Ph.D. Rutgers 1979. Nondisabled-disabled wage differentials: An empirical test for discrimination.

Francis W. Horvath, Ph.D. North Carolina (Chapel Hill) 1980. An economic analysis of migration in the United States: Repeat movements and families.

Vincent J. Hotz, Ph.D. Wisconsin (Madison) 1980. A theoretical and empirical model of fertility and married women's allocation of time over the life cycle.

Janet Johnson, Ph.D. Wisconsin (Madison) 1980. A job search model of household unemployment behavior.

David E. Keefe, Ph.D. California (Berkeley) 1980. The fertility of women on welfare.

Randall G. Kesselring, Ph.D. Oklahoma 1980. A case analysis of Section B(a) (2) of the Taft-Hartley Act, 1950-1974.

George A. Kettner, Ph.D. California (Los Angeles) 1980. Human capital investment in alternative college majors.

Choongsoo Kim, Ph.D. Pennsylvania 1979. Three essays on the youth labor market: Time-series studies of military enlistment, school enrollment, and unemployment.

Charles T. Koeller, Ph.D. Rutgers 1979. A bargaining model of American wage determination.

James Lambrinos, Ph.D. Rutgers 1979. Health, income maintenance, and labor supply.

Mark D. Lange, Ph.D. Iowa State 1979. An economic analysis of time allocation and capital-labor ratios in household production of farm families in Iowa.

Samuel A. Laposta, Ph.D. Pittsburgh 1979. The effect of imperfections in experience rating on seasonal employment.

Donald A. Larson, Ph.D. Washington (St. Louis) 1979. The components of nonmarket time and female labor supply patterns.

JOHN J. LAWLER, Ph.D. California (Berkeley) 1980. Interdependent wage adjustment in contract construction: The impact of intra-union power and market uncertainty.

KAREN LEPPEL, Ph.D. Princeton 1980. The relations among child quality, family structure, and the value of the mother's time.

R. RODMAN LUDLOW, Ph.D. Purdue 1979. A human capital approach to first, repeat, and return migration decisions of young men in the United States.

VENKATA B. N. S. MADDURI, Ph.D. Alberta 1979. Economic analysis of fertility behavior in Canada: An examination of 1971 Census data.

MARK MAIER, Ph.D. New School 1980. Municipal labor relations in New York City, 1954-73.

WILLIAM L. MOORE, Ph.D. California (Berkeley) 1979. Determinants and outcomes of departmental power: A two-campus study.

LAURENCE C. MORSE, Ph.D. Princeton 1980. Essays on youth labor market problems.

STEVEN C. MYERS, Ph.D. Ohio State 1979. Working in college: Risk or return?

MICHAEL K. NAKADA, Ph.D. California (Santa Barbara) 1979. A theoretical and empirical analysis of the labor supply of wives with husbands employed either full time or part time.

MICHAEL B. ORMISTON, Ph.D. Johns Hopkins 1980. On life cycle models with endogenous labor-leisure choices and skills.

JAMES A. ORR, Ph.D. Pennsylvania 1979. A model of unemployment in local labor markets with an application to discretionary fiscal programs.

JEAN T. OWEN, Ph.D. Boston 1980. Interregional labor migration in French indicative planning.

ALVARO PACHON, Ph.D. Harvard 1980. Integrating human capital in the National Accounts: Conceptual framework and empirical implementation.

HAMES E. PEARCE, Ph.D. California (Los Angeles) 1980. Implicit contracting, trade unionization, and the response to demand variation in U.S. manufacturing.

WOLFGANG POLLAN, Ph.D. California (Berkeley) 1980. Adjustment costs, labor turnover, and employment changes.

CHARLES POULTON-CALLAHAN, Ph.D. Illinois 1979. An analysis of the labor market experience of young women.

FRANCISCO L. RIVERA-BATIZ, Ph.D. Massachusetts Institute of Technology 1980. Three essays on the theory of labor migration.

JAMES R. ROOK, Ph.D. North Carolina State 1979. Labor skill composition, elasticities of substitution and industrial migration: The case of cotton textiles.

HOWARD R. ROSENBERG, Ph.D. California (Berkeley) 1980. The part-time MBA: Impacts on the mid-career development of managers.

BAHK SANG, Ph.D. New School 1980. Insured unemployment and availability of unemployment insurance benefits in the United States.

MORTON O. SCHAPIRO, Ph.D. Pennsylvania 1979. A general equilibrium model of northern farm settlement.

LEONARD A. SCHLESINGER, D.B.A. Harvard 1979. Supervisory roles in participative work systems.

PHILIP E. SCHOECH, Ph.D. Wisconsin (Madison) 1980. Measuring economic performance in the Netherlands, 1951-76.

JANET L. SCHOLL, Ph.D. Michigan State 1980. Transition from unemployment: Outcome and duration of individual spells.

RICHARD D. SCOTT, Ph.D. British Columbia 1979. Schooling, experience, hours of work, and earnings in Canada.

STEVEN H. SHEINGOLD, Ph.D. Pennsylvania State 1979. Specific capital, implicit contracts, and the layoff process: An application to evaluating the Trade Adjustment Assistance Program.

DAVID A. SMITH, Ph.D. Massachusetts Institute of Technology 1980. Labor market participation, turnover, and the Gary Income Maintenance Experiment.

SUSAN C. STEPHENSON, Ph.D. Iowa State 1979. The impact of functional limitations on labor supply.

CAROLINE SWARTZ, Ph.D. Duke 1979. Human capital, labor force participation, and the extended family in urban Kenya.

WARREN T. TREPETA, Ph.D. Brown 1980. Reliability and the racial differentiation in income stability.

THAM V. TRUON, Ph.D. Kentucky 1979. Less-developed countries' industrial labor absorptive capacity and economic integration: Central American Common Market as a case study.

DONALD WALDMAN, Ph.D. Wisconsin (Madison) 1979. Time allocation of young men.

ALLEN L. WEBSTER, Ph.D. Florida State 1979. An analysis of public service employment programs under the Comprehensive Employment and Training Act in the State of Florida.

DOUGLAS A. WION, Ph.D. Pennsylvania State 1980. Increases in the production of working wives and the distribution of family earnings.

HUSSEIN YAHYA, Ph.D. Oklahoma State 1980. Human capital migration from labor-rich Arab states to oil-rich Arab states and the consequences for the Jordanian economy.

DAVID W. YANG, Ph.D. Nebraska (Lincoln) 1979. Interindustry analysis applied to racial groups in Pine Bluff, Arkansas.

UGUR YUCELT, Ph.D. New School 1980. Absenteeism and its causes in the industrial sector of developing nations: The Turkish case.

Welfare Programs; Consumer Economics; Urban and Regional Economics

THEODORE H. AHLERS, Ph.D. Tufts 1979. A microeconomic analysis of rural-urban migration in Haiti.

ROBERT B. AVERY, Ph.D. Wisconsin (Madison) 1979. The role of uncertainty in the savings decisions of low-income families.

JOHN W. BAER, D.A. Catholic 1980. Specific instructional objectives in the principles of economic courses in the Maryland state community colleges.

ROBERT A. BATTS, Ph.D. Kentucky 1979. Discrimination: Its costs and effects on minority groups in the South.

JOHN E. BELKNAP, Ph.D. Wisconsin (Madison) 1979. Human capital accumulation by rural families with respect to variation in family income.

MARION G. BLALOCK, Ph.D. Oklahoma State 1980. An examination of the extent and impact of industrialization in rural Oklahoma communities.

THOMAS P. BOEHM, Ph.D. Washington (St. Louis) 1979. The interrelationship between housing choice and expected utility.

DEBORAH A. J. BROWN, Ph.D. Purdue 1979. Application of principal-agent theory to health care costs.

HAROLD L. BUNCE, Ph.D. Indiana 1980. An econometric study of state and local government expenditures.

OH-HYUN CHANG, Ph.D. Wisconsin (Madison) 1979. Impact of permanent income, prices, and sociodemographic characteristics on household expenditure patterns in the United States, 1960-61.

THOMAS CHESTER, Ph.D. Michigan State 1979. Minimum mandatory sentences and plea bargaining: An economic perspective.

RICHARD COE, Ph.D. Michigan 1979. Participation in the Food Stamp Program.

H. MICHAEL COINER, Ph.D. Yale 1979. Three essays in the economics of higher education.

MICHAEL J. DANIELS, Ph.D. Georgia State 1979. An econometric model of a region's financial sector.

EDWARD E. DORR, Ph.D. North Carolina State 1979. Economies of scale in high school construction and operation.

RADCLIFFE EDMONDS, Ph.D. Michigan 1979. Residential land values and the willingness to pay for central city access in metropolitan Tokyo.

EDWARD A. FAGERLUND, Ph.D. Minnesota 1979. The urban settlement pattern and the cost of providing wastewater disposal service.

DANIEL R. FEENBERG, Ph.D. Princeton 1980. Estimating the demand for water quality from market data.

MERTON D. FINKLER, Ph.D. Minnesota 1979. The influence on municipal fiscal policy on single family residential property value.

LOUIS G. FORTIS, Ph.D. Massachusetts 1980. A community-oriented strategy for the revitalization of the urban centers of the Northeast and upper Midwest: The tradeoff between the short-run micro-economic decisions and the long-term macro-economic imperative.

DAVID L. FRANKLIN, Ph.D. North Carolina State 1979. The derived demand for calories in households with malnourished children in Cali, Colombia.

THOMAS W. GRANNEMANN, Ph.D. Northwestern 1979. The demand for publicly financed medical care: The role of interdependent preferences.

AHMAD HAMZAH, Ph.D. Tufts 1980. The political economy of oil in Brunei.

ROGER HITCHNER, Ph.D. Chicago 1979. An economic analysis of causes of obesity in U.S. adults.

JOSEPH D. HURD II, Ph.D. California (Berkeley) 1979. Enrollment and cost effects of financial aid plans for higher education.

HELEN H. JENSEN, Ph.D. Wisconsin (Madison) 1980. Health insurance in rural areas.

EMMANUEL Y. JIMENEZ, Ph.D. Brown 1980. The economics of self-help housing: Theory and evidence in the development setting.

JEFFREY M. KALLIS, Ph.D. Pittsburgh 1979. The economics of crime and deterrence: A regulatory view of public policy.

DONALD L. LERMAN, Ph.D. Wisconsin (Madison) 1980. Economics of school closings.

PATRICIA K. LITTLE, Ph.D. Washington (St. Louis) 1980. A theory of the multicentered city.

ANA MARIA T. LOMPERIS, Ph.D. North Carolina (Chapel Hill) 1979. An economic analysis of the household's production of preschool child nutritional health.

ANTHONY L. LOVISCEK, Ph.D. West Virginia 1980. Interindustry linkages and industrial clusters: A comparative analysis.

MICHAEL D. MAHER, Ph.D. Rice 1980. The economics of urban stormwater management.

GALE D. MERSETH, D.B.A. Harvard 1979. Regulation and income redistribution: The Massachusetts experience.

WILLIAM J. MILNE, Ph.D. Pennsylvania 1980. A multiregional econometric model of the United States: An application to the Northeast.

LAURENCE A. MINERS, Ph.D. North Carolina (Chapel Hill) 1979. The family's demand for health: A rural investigation.

PETER J. MORGAN, Ph.D. Yale 1979. The economics of joint vs. separate treatment of domestic and industrial wastewaters.

MICHAEL A. MORRISEY, Ph.D. Washington 1979. The effects of Medicaid reimbursement on nursing home services.

NADIM MUNLA, Ph.D. Southern California 1980. Interaction between the public and private sectors in regional I-O models.

THOMAS O'GRADY, Ph.D. California (Berkeley) 1980. Demographic changes and their effects on urban household formations and locations.

TAE KYU PARK, Ph.D. Indiana 1980. The income redistributive effects of national health insurance.

MARTHA S. PEYTON, Ph.D. Fordham 1980. The effect of neighborhood location on mortgage risk.

GEORGE M. RADAKOVIC, Ph.D. Pittsburgh 1979. The impact of market power on metropolitan areas.

STEPHEN W. RITTERBUSH, Ph.D. Tufts 1980. The economics and politics of natural resources in southern Africa: A case study of Namibia.

JENNIFER A. ROBACK, Ph.D. Rochester 1980. The value of local urban amenities: Theory and measurement.

UWE SCHUBERT, Ph.D. California (San Diego) 1979. Environmental quality and urban land markets.

ROBERT M. SCHWAB, Ph.D. Johns Hopkins 1980. Inflation expectations and the demand for housing.

RONALD J. SEPANIK, Ph.D. Wisconsin (Madison) 1979. Income tax advantage and owner-occupancy: A new answer to an old question.

SHAHIN SHAHIDSALESS, Ph.D. Wisconsin (Madison) 1979. The impact of socioeconomic and location factors on the employment outcome of manufacturing expansion in the nonmetropolitan United States.

JODY L. SINDELAR, Ph.D. Stanford 1980. Differential utilization of medical care by sex: A theoretical and empirical study.

ARTHUR SNOW, Ph.D. Wisconsin (Madison) 1979. Consumer information and consumer protection.

STEVEN D. SODERLIND, Ph.D. Massachusetts 1980. Norwegian regional economic development policy.

ERNEST B. STOKES, Ph.D. Queen's 1979. Housing markets in Canada: An econometric analysis using regional panel data.

SHIRLEY V. SVORNY, Ph.D. California (Los Angeles) 1979. Foreign-trained physicians and health care in the United States.

ELLEN P. TANZER, Ph.D. Wisconsin (Madison) 1979. The effect on housing quality of replacing property by site-value taxation.

BOON N. TAY, Ph.D. Hawaii 1980. Demand for subsidized housing in Hawaii.

JOHN P. TILLMAN, Ph.D. Nebraska (Lincoln) 1979. A descriptive model for comprehensive regional growth analysis with an application to the economics of Nebraska and the United States, 1964-76.

RAYMOND F. TURNER, Ph.D. Catholic 1980. The two-year college as an institution for retraining four-year college graduates: An economic analysis with case studies.

ERIC N. VAN DE VERG, Ph.D. Claremont 1980. Housing prices in Norfolk neighborhoods.

JAMES R. WARD, Ph.D. Michigan State 1979. The relationship between client characteristics and those nonmedical home care services and costs associated with the long-term care of the impaired aged in Michigan.

DOUGLAS S. WEST, Ph.D. British Columbia 1979. Market preemption as a barrier to entry in a growing, spatially extended market.

MARC R. WIITALA, Ph.D. Nebraska (Lincoln) 1979. A theoretical and empirical analysis of two gravity-deterministic industrial location models.

MARIANNE M. WOLD, Ph.D. Johns Hopkins 1980. Public investment in transportation, expected transportation costs, and the urban housing market.

JOHN R. WOLFE, Ph.D. Pennsylvania 1979. The impact of parental resources on childhood IQ.

THOMAS J. ZLATOPER, Ph.D. Northwestern 1980. Conditional logit estimates of grain elevator shipment models with fixed and endogenous transport rates.

THE AMERICAN ECONOMIC REVIEW

VOLUME LXX

BOARD OF EDITORS

Albert Ando
Pranab K. Bardhan
Peter A. Diamond
W. Erwin Diewert
Rudiger Dornbusch
Herschel I. Grossman
Peter W. Howitt
Anne O. Krueger
William H. Oakland
Michel Parkin
Roy Radner
Richard R. Roll
Nancy L. Schwartz
James P. Smith
A. Michael Spence
William S. Vickrey
Robert D. Willig
S. Y. Wu

MANAGING EDITOR

George H. Borts

THE AMERICAN ECONOMIC ASSOCIATION

Executive Office: Nashville, Tennessee

Editorial Office: Brown University, Providence, Rhode Island

Copyright 1980

All Rights Reserved

AMERICAN ECONOMIC ASSOCIATION

CONTENTS OF ARTICLES AND SHORTER PAPERS

R. M. Solow: On Theories of Unemployment . . . 1

V. Tanzi: Inflationary Expectations, Economic Activity, Taxes and Interest Rates 12

M. Okuno, A. Postlewaite, and J. Roberts: Oligopoly and Competition in Large Markets . 22

B. M. Friedman: Price Inflation, Portfolio Choice, and Nominal Interest Rates 32

W. L. Hansen, H. B. Newburger, F. J. Schroeder, D. C. Stapleton, and D. J. YoungDay: Forecasting the Market for New Ph.D. Economists . 49

S. D. Braithwait: The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes 64

J. Sachs: The Changing Cyclical Behavior of Wages and Prices: 1890–1976 78

E. P. Lazear and R. T. Michael: Family Size and the Distribution of Real Per Capita Income . . 91

P. Chinloy: Sources of Quality Change in Labor Input 108

D. W. Caves and E. L. Feige: Efficient Foreign Exchange Markets and the Monetary Approach to Exchange-Rate Determination 120

J. H. Wood: Metzler on Classical Interest Theory . 135

D. J. Mullineaux: Inflation Expectations and Money Growth in the United States 149

R. Stevenson: Measuring Technological Bias . . . 162

K. D. Boyer: Queuing Analysis and Value of Service Pricing in the Trucking Industry: Comment 174

A. S. DeVany and T. R. Saving: Reply 181

J. S. Akin and I. Garfinkel: The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Comment 186

F. Welch: Reply 192

C. Link, E. Ratledge, and K. Lewis: Reply 196

D. G. Davies: Measurement of Tax Progressivity: Comment 204

E. C. Kienzle: Comment 208

D. B. Suits: Reply 211

J. Paroush and N. Kahana: Price Uncertainty and the Cooperative Firm 212

C. Lim: The Ranking of Behavioral Modes of the Firm Facing Uncertain Demand 217

D. Friedman: Many, Few, One: Social Harmony and the Shrunken Choice Set 225

L. Young: Distributional Neutrality and Optimal Commodity Taxation: Comment 233

D. E. Wildasin: Reply 237

M. S. Ahluwalia, J. H. Duloy, G. Pyatt, and T. N. Srinivasan: Who Benefits from Economic Development?: Comment 242

P. Beckerman and D. Coes: Comment 246

A. Fishlow: Comment 250

G. S. Fields: Reply 257

R. L. Jacobs and R. A. Jones: Price Expectations in the United States, 1947–75 269

R. N. Batra and R. Ramachandran: Multinational Firms and the Theory of International Trade and Investment 278

R. Findlay: The Terms of Trade and Equilibrium Growth in the World Economy 291

J. E. Harris: Taxing Tar and Nicotine 300

A. Deaton and J. Muellbauer: An Almost Ideal Demand System 312

S. A. Hoenack and W. C. Weiler: A Structural Model of Murder Behavior and the Criminal Justice System 327

H. S. Burness, W. D. Montgomery, and J. P. Quirk: Capital Contracting and the Regulated Firm 342

G. J. Duncan and F. P. Stafford: Do Union Members Receive Compensating Wage Differentials? 355

M. J. Bailey, M. Olson, and P. Wonnacott: The Marginal Utility of Income Does Not Increase: Borrowing, Lending, and Friedman-Savage Gambles 372

D. A. Starrett: On the Method of Taxation and the Provision of Local Public Goods 380

S. J. Grossman and J. E. Stiglitz: On the Impossibility of Informationally Efficient Markets . . . 393

R. T. Froyen and R. N. Waud: Further International Evidence on Output-Inflation Tradeoffs 409

D. W. Caves and L. R. Christensen: Global Properties of Flexible Functional Forms 422

M. Farrell: International Impact of U.S. Money Supply: The Case of Mexico 433

S. Oster: The Optimal Order for Submitting Manuscripts 444

A. Randall and J. R. Stoll: Consumer's Surplus in Commodity Space 449

S. P. Das: On the Effect of Rate-of-Return Regulation under Uncertainty 456

R. J. Staaf and E. G. West: Limits on Public Provision of Private Goods: (Comment) 461

A. T. Denzau and R. J. Mackay: Reply and Further Results 466

R. Halvorsen and R. Palmquist: The Interpretation of Dummy Variables in Semilogarithmic Equations 474

N. L. Jacob and A. N. Page: Production, Information Costs, and Economic Organization: The Buyer Monitoring Case 476

J. P. Formby and T. G. Seaks: Paglin's Gini Measure of Inequality: A Modification 479

R. M. Bradburd: Conglomerate Power without Market Power: The Effects of Conglomeration on a Risk-Averse Quantity-Adjusting Firm 483

G. Philpotts: A Note on the Representation of Preferences in the Lindahl-Johansen Diagram 488

J. J. Merrick, Jr.: Measuring the Expected Real Rate of Interest—An Exploration of Macro-Economic Alternatives: Comment 493

J. W. Elliott: Reply and Further Thoughts 498

T. S. McCaleb: Excess Burden, Benefit Taxation, and Efficiency in Public Expenditure 501

E. Helpman and D. Pines: Optimal Public Investment and Dispersion Policy in a System of Open Cities 507

H. Akiba: Gains from Trade under Uncertainty: Further Comment 515

J. P. Hughes: The Comparative Statics of the Competitive Increasing-Cost Industry 518

L. Young: Tariffs vs. Quotas under Uncertainty: An Extension 522

S. P. Das: Further Results on Input Choices under Uncertain Demand 528

W. A. Lewis: The Slowing Down of the Engine of Growth 555

J. L. Stein: The Dynamics of Spot and Forward Prices in an Efficient Foreign Exchange Market with Rational Expectations 565

V. L. Smith: Experiments with a Decentralized Mechanism for Public Good Decisions 584

M. Aoki: A Model of the Firm as a Stockholder-Employee Cooperative Game 600

R. F. Lucas: Tariffs, Nontraded Goods, and the Optimal Stabilization Policy 611

D. O. Parsons: Unemployment, the Allocation of Labor, and Optimal Government Intervention 626

M. Feldstein: Fiscal Policies, Inflation, and Capital Formation 636

H. R. Varian: A Model of Sales 651

P. L. Cheng and R. R. Grauer: An Alternative Test of the Capital Asset Pricing Model 660

R. W. Fogel and S. L. Engerman: Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Reply 672

D. L. Brito and W. H. Oakland: On the Monopolistic Provision of Excludable Public Goods 691

J. Eaton and H. S. Rosen: Taxation, Human Capital, and Uncertainty 705

E. Helpman and A. Razin: Efficient Protection under Uncertainty 716

J. K. Stephens: Unemployment, Inflation, and Monetarism—A Further Analysis: Comment 732

R. Van Order: Reply 738

H. Yoshikawa: On the "q" Theory of Investment 739

Y.-K. Ng: Optimal Corrective Taxes or Subsidies when Revenue Raising Imposes an Excess Burden 744

L. Samuelson: The Golden Rule with Endogenous Labor Participation Rate 752

S. I. Cohen: Incentives and the Choice of Optimal Plans 756

M. K. Salemi: The Forward Exchange Rate, Expectations, and the Demand for Money—The German Hyperinflation: Comment 763

J. A. Frenkel: Reply 771

R. Driskill: Exchange Rate Dynamics, Portfolio Balance, and Relative Prices 776

W. M. Crain and A. Zardkoohi: X-Inefficiency and Nonpecuniary Rewards in a Rent-Seeking Society: A Neglected Issue in the Property Rights Theory of the Firm 784

D. N. Stevens and J. E. Foster: Comparing Utility Functions in Efficiency Terms: Comment 793

B. A. Weisbrod: Reply 795

F. H. Stephen: Bank Credit and the Labor-Managed Firm: Comment 796

E. G. Furubotn: Reply 800

J. S. Hekman: Income, Labor Supply, and Urban Residence 805

H. Herberg and M. C. Kemp: In Defense of Some "Paradoxes" of Trade Theory 812

J. P. Neary: This Side of Paradox, or, in Defense of the Correspondence Principle: A Reply to Herberg and Kemp 815

P. Howitt and D. Patinkin: Utility Function Transformations and Money Illusion: Comment 819

R. Dusansky: Reply and Further Results 823

P. Howitt and D. Patinkin: A Further Note 826

I. M. McDonald: On the Comparison of the Stability Implications of Marshallian and Walrasian Adjustment Schemes: Note 829

M. Feldstein: Inflation and the Stock Market 839

L. W. Lee: A Theory of Just Regulation 848

E. Mansfield: Basic Research and Productivity Increase in Manufacturing 863

R. E. Verrecchia: Consensus Beliefs, Information Acquisition, and Market Information Efficiency 874

G. Akerlof and B. G. M. Main: Unemployment Spells and Unemployment Experience 885

J. V. Henderson: Community Development: The Effects of Growth and Uncertainty 894

D. O. Parsons: Racial Trends in Male Labor Force Participation 911

C. Menezes, C. Geiss, and J. Tressler: Increasing Downside Risk 921

J. S. Chipman, and J. C. Moore: Compensating Variation, Consumer's Surplus, and Welfare 933

P. Krugman: Scale Economics, Product Differentiation, and the Pattern of Trade 950

R. Dornbusch and S. Fischer: Exchange Rates and the Current Account 960

J. A. Hanson: The Short-Run Relation Between Growth and Inflation in Latin America: A Quasi-Rational or Consistent Expectations Approach 972

G. V. Jump: Interest Rates, Inflation Expectations, and Spurious Elements in Measured Real Income and Saving 990

R. E. Lucas, Jr.: Two Illustrations of the Quantity Theory of Money 1005

K. Pertz: Reswitching, Wicksell Effects, and the Neoclassical Production Function: Note 1015

J. H. McCulloch: The Ban on Indexed Bonds, 1933–77 1018

M. D. Levi and J. H. Makin: Inflation Uncertainty and the Phillips Curve: Some EmpiricalEvidence 1022

E. O. Olsen: Pareto-Desirable Redistribution in Kind: Comment 1028

G. Brennan and C. Walsh: Reply 1032

A. H. Barnett: The Pigouvian Tax Rule under Monopoly 1037

A. Hurter, J. S. Martinich, and E. R. Venta: A Note on the Separability of Production and Location 1042

D. R. Capozza and R. Van Order: Unique Equilibria, Pure Profits and Efficiency in Location Models 1046

E. Silberberg: Harold Hotelling and Marginal Cost Pricing 1054

V. A. Aivazian and J. L. Callen: Uncertain Externalities, Liability Rules, and Resource Allocation: Comment 1058

P. H. Greenwood and C. A. Ingene: Reply 1060

L. Eeckhoudt and P. Hansen: Minimum and Maximum Prices, Uncertainty, and the Theory of the Competitive Firm 1064

B. B. Gibson: Estimating Demand Elasticities for Public Goods from Survey Data 1069

R. Dardis: The Value of a Life: New Evidence from the Marketplace 1077

S. Perrakis: Factor-Price Uncertainty with Variable Proportions: Note 1083

F. D. Arditti and H. Levy: A Model of the Parallel Team Strategy in Product Development . . . 1089

B. L. Benson: Löschian Competition under Alternative Demand Conditions 1098

CONTENTS OF THE PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

T. Scitovsky: Can Capitalism Survive?—An Old Question in a New Setting 1

Appraisals of Post-Keynesian Economics

L. Tarshis: Post-Keynesian Economics: A Promise that Bounced? 10

J. L. Yellen: On Keynesian Economics and the Economics of Post-Keynesians 15

J. R. Crotty: Post-Keynesian Economic Theory: An Overview and Evaluation 20

Discussion by P. Kenyon; G. C. Harcourt 26

Studies of Teaching and Learning in Economics

J. J. Siegfried and G. H. Sweeney: Bias in Economics Education Research from Random and Voluntary Selection into Experimental and Control Groups 29

W. E. Becker, Jr. and M. J. Morey: Pooled Cross-Section Time-Series Evaluation: Source, Result, and Correction of Serially Correlated Errors 35

M. K. Salemi and G. E. Tauchen: Guessing and the Error Structure of Learning Models . . . 41

Discussion by G. Chamberlain 47

Innovation, Technological Progress, and Research and Development

G. C. Eads: Regulation and Technical Change: Some Largely Unexplored Influences 50

N. E. Terleckyj: What Do *R&D* Numbers Tell Us About Technological Change? 55

R. R. Nelson: Production Sets, Technological Knowledge, and *R&D*: Fragile and Overworked Constructs for Analysis of Productivity 62

Discussion by R. Piekarz 68

The Effects of Fiscal Policies on the Distribution of Income and Wealth

M. A. King: How Effective Have Fiscal Policies Been in Changing the Distribution of Income and Wealth? 72

R. M. Bird: Income Redistribution Through the Fiscal System: The Limits of Knowledge . . . 77

L. C. Thurow: The Indirect Incidence of Government Expenditures 82

Discussion by R. H. Haveman; P. Mieszkowski . . 88

Long-Term Growth in Third-World Economies

L. G. Reynolds: Economic Development in Historical Perspective 91

A. Heston and R. Summers: Comparative Indian Economic Growth: 1870–1970 96

A. Fishlow: Brazilian Development in Long-Term Perspective 102

Discussion by M. D. Morris; S. A. Morley 109

The Recoil from Welfare Capitalism: Political and Sociological Perspectives

A. O. Hirschman: The Welfare State in Trouble: Systemic Crisis or Growing Pains? 113

I. Katznelson: Accounts of the Welfare State and the New Mood 117

Equality Incentives and Economic Policy

R. W. Palmer: Equality, Incentives, and Economic Power 123

M. Reich: The Persistence of Racial Inequality in Urban Areas and Industries, 1950–70 128

S. Wilson, D. Steinberg, and J. C. Kulik: Guaranteed Employment, Work Incentives, and Welfare Reform: Insight from the Work Equity Project 132

L. Anderson: Rates of Return to Human Capital: A Test Using El Salvador Data 138

The Political Economy of National Health Insurance

A. C. Enthoven: How Interested Groups have Responded to a Proposal for Economic Competition in Health Services 142

U. E. Reinhardt: Health Insurance and Cost-Containment Policies: The Experience Abroad 149

The Government and Capital Formation

R. E. Hall: Stabilization Policy and Capital Formation 157

M. J. Boskin and J. B. Shoven: Issues in the Taxation of Capital Income in the United States 164

Taxation, Labor Supply, and Saving

H. S. Rosen: What is Labor Supply and Do Taxes Affect It? 171

G. M. von Furstenberg: Private Saving 177

M. Feldstein: Tax Rules and the Mismanagement of Monetary Policy 182

Discussion by M. Kosters; J. B. Shoven; A. S. Blinder 187

Consequences of the Growth of the Two-Earner Family

J. F. Madden: Urban Land Use and the Growth in Two-Earner Households 191

J. A. Matthaei: Consequences of the Rise of the Two-Earner Family: The Breakdown of the Sexual Division of Labor 198

E. P. Lazear and R. T. Michael: Real Income Equivalence Among One-Earner and Two-Earner Families 203

D. C. Quinlan and J. A. Shackelford: Labor Force Participation Rates of Women and the Rise of the Two-Earner Family 209

Discussion by M. O. Keating 213

A General View of Capital Formation and Economic Growth

R. M. Coen and B. G. Hickman: Investment and Growth in an Econometric Model of the United States 214

E. F. Denison: The Contribution of Capital to Economic Growth 220

R. Eisner: Total Income, Total Investment, and Growth 225

Discussion by J. W. Kendrick; O. Havrylyshyn . . 232

Interwar Macroeconomics from the Perspective of the 1970's

J. A. Frenkel: Exchange Rates, Prices, and Money: Lessons from the 1920's 235

R. J. Gordon: A Consistent Characterization of a Near-Century of Price Behavior 243

C. A. Sims: Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered . 250

Discussion by A. H. Meltzer 258

Recent Developments in the Economic Theory of Index Numbers

W. E. Diewert: Capital and the Theory of Productivity Measurement 260

D. W. Jorgenson, L. J. Lau, and T. M. Stoker: Welfare Comparison under Exact Aggregation 268

R. A. Pollak: Group Cost-of-Living Indexes 273

Discussion by L. R. Christensen 279

Changes in Trade Shares and Economic Growth

H. B. Chenery: Interactions Between Industralization and Exports 281

A. O. Krueger: Trade Policy as an Input to Development 288

C. P. Kindleberger: Government Policies and Changing Shares in World Trade 293

Discussion by C. F. Diaz Alejandro 299

Regulatory Intervention in Historical Perspective

P. Temin: Regulation and the Choice of Prescription Drugs 301

T. S. Ulen: The Market for Regulation: The ICC from 1887 to 1920 306

J. C. Panzar: Regulation, Deregulation, and Economic Efficiency: The Case of the CAB . . . 311

Discussion by R. G. Noll; E. W. Hawley; M. E. Levine 316

Empirical Studies of the Rate of Return to Capital

D. M. Holland and S. C. Myers: Profitability and Capital Costs for Manufacturing Corporations and All Nonfinancial Corporations . . . 320

B. M. Fraumeni and D. W. Jorgenson: Rates of Return by Industrial Sector in the United States, 1948–76 326

A. C. Harberger: Vignettes on the World Capital Market 331

Discussion by B. S. Bernanke and J. I. Bulow . . . 338

Current Retardation in U. S. Productivity Growth

A. Rees: Improving Productivity Measurement . . 340

Z. Griliches: *R&D* and the Productivity Slowdown 343

M. I. Nadiri: Sectoral Productivity Slowdown . . . 349

Discussion by J. A. Mark; E. F. Denison 353

Borderlines of Law and Economic Theory

B. Klein: Transaction Cost Determinants of "Unfair" Contractual Arrangements 356

A. M. Polinsky: Strict Liability vs. Negligence in a Market Setting 363

The Evolving World Dollar Standard

S. P. Magee and R. K. S. Rao: Vehicle and Nonvehicle Currencies in International Trade . . . 368

J. A. Frenkel and M. L. Mussa: The Efficiency of the Foreign Exchange Market and Measures of Turbulence 374

R. I. McKinnon: Dollar Stabilization and American Monetary Policy 382

Changing Role in the Market in Utilities Regulation

H. M. Trebing: Structural Change and Regulatory Reform in the Utilities Industries 388

W. H. Melody: Radio Spectrum Allocation: Role of the Market 393

T. K. Standish: State Initiatives in State-Federal Relations 398

Discussion by D. W. Smythe 403

The Decline of American Cities: Self-Correcting and Self-Aggravating Forces

W. Alonso: Population as a System in Regional Development 405

K. L. Bradbury, A. Downs, and K. A. Small: Some Dynamics of Central City-Suburban Interactions 410

M. A. Stegman and D. W. Rasmussen: Neighborhood Stability in Changing Cities 415

Discussion by H. A. Garn; W. H. Frey; J. R. Pack 420

◆

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS

Ahluwalia, M. S. 242
Aivazian, V. A. 1058
Akerlof, G. 885
Akiba, H. 515
Akin, J. S. 186
Aoki, M. 600
Arditti, F. D. 1089
Bailey, M. J. 372
Barnett, A. H. 1037
Batra, R. N. 278
Beckerman, P. 246
Benson, B. L. 1098
Boyer, K. D. 174
Bradburd, R. 483
Braithwait, S. D. 64
Brennan, G. 1032
Brito, D. L. 691
Burness, H. S. 342
Callen, J. L. 1058
Capozza, D. R. 1046
Caves, D. W. 120, 422
Cheng, P. L. 660
Chinloy, P. 108
Chipman, J. S. 933
Christensen, L. R. 422
Coes, D. 246
Cohen, S. I. 756
Crain, W. M. 784
Dardis, R. 1077
Das, S. P. 456, 528
Davies, D. G. 204
Deaton, A. 312
Denzau, A. T. 466
DeVany, A. S. 181
Dornbusch, R. 960
Driskill, R. 776
Duloy, J. H. 242
Duncan, G. J. 355
Dusansky, R. 823
Eaton, J. 705
Eeckhoudt, L. 1064
Elliott, J. W. 498
Engerman, S. L. 672
Farrell, M. 433
Feige, E. L. 120
Feldstein, M. 636, 839
Fields, G. S. 257
Findlay, R. 291
Fischer, S. 960
Fishlow, A. 250
Fogel, R. W. 672
Formby, J. P. 479
Foster, J. E. 793
Frenkel, J. A. 771
Friedman, B. M. 32
Friedman, D. 225
Froyen, R. T. 409
Furubotn, E. G. 800
Garfinkel, I. 186
Geiss, C. 921
Gibson, B. B. 1069
Grauer, R. R. 660
Greenwood, P. H. 1060
Grossman, S. J. 393
Halvorsen, R. 474
Hansen, P. 1064
Hansen, W. L. 49
Hanson, J. A. 972
Harris, J. E. 300
Hekman, J. S. 805
Helpman, E. 507, 716
Henderson, J. V. 894
Herberg, H. 812
Hoenack, S. A. 327
Howitt, P. 819, 826
Hughes, J. P. 518
Hurter, A. P. 1042
Ingene, C. A. 1060
Jacob, N. L. 467
Jacobs, R. L. 269
Jones, R. A. 269
Jump, G. V. 990
Kahana, N. 212
Kemp, M. C. 812
Kienzle, E. C. 208
Krugman, P. 950
Lazear, E. P. 91
Lee, L. W. 848
Levi, M. D. 1022
Levy, H. 1089
Lewis, K. 196
Lewis, W. A. 555
Lim, C. 217
Link, C. 196
Lucas, R. E. Jr., 1005
Lucas, R. F. 611
McCaleb, T. S. 501
McCulloch, J. H. 1018
McDonald, I. 829
Mackay, R. J. 466
Main, B. G. M. 885
Makin, J. H. 1022
Mansfield, E. 863
Martinich, J. S. 1042
Menezes, C. 921
Merrick, J. J. Jr., 493
Michael, R. T. 91
Montgomery, W. D. 342
Moore, J. C. 933
Muellbauer, J. 312
Mullineaux, D. J. 149
Neary, J. P. 815
Newburger, H. B. 49
Ng, Y.-K. 744
Oakland, W. H. 691
Okuno, M. 22

Olsen, E. O. 1028
Olson, M. 372
Oster, S. 444
Page, A. N. 476
Palmquist, R. 474
Paroush, J. 212
Parsons, D. O. 626, 911
Patinkin, D. 819, 826
Perrakis, S. 1083
Pertz, K. 1015
Philpotts, G. 488
Pines, D. 507
Postlewaite, A. 22
Pyatt, G. 242
Quirk, J. P. 342
Ramachandran, R. 278
Randall, A. 449
Ratledge, E. 196
Razin, A. 716
Roberts, J. 22
Rosen, H. S. 705
Sachs, J. 78
Salemi, M. K. 763
Samuelson, L. 752
Saving, T. R. 181
Schroeder, F. J. 49
Seaks, T. G. 479
Silberberg, E. 1054
Smith, V. L. 584
Solow, R. M. 1
Srinivasan, T. N. 242
Staaf, R. J. 461
Stafford, F. P. 355
Stapleton, D. C. 49
Starrett, D. A. 380
Stein, J. L. 565
Stephen, F. H. 796
Stephens, J. K. 732
Stevens, D. N. 793
Stevenson, R. 162
Stiglitz, J. E. 393
Stoll, J. R. 449
Suits, D. B. 211
Tanzi, V. 12
Tressler, J. 921
Van Order, R. 738, 1046
Varian, H. R. 651
Venta, E. R. 1042
Verrecchia, R. E. 874
Walsh, C. 1032
Waud, R. N. 409
Weiler, W. C. 327
Weisbrod, B. A. 795
Welch, F. 192
West, E. G. 461
Wildasin, D. E. 237
Wonnacott, P. 372
Wood, J. H. 135
Yoshikawa, H. 739
Young, L. 233, 522
YoungDay, D. J. 49
Zardkoohi, A. 784

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Alonso, W. 405
Anderson, L. 138
Becker, W. E. Jr. 35
Bernanke, B. S. 338
Bird, R. M. 77
Blinder, A. S. 189
Boskin, M. J. 164
Bradbury, K. L. 410
Bulow, J. I. 338
Chamberlain, G. 47
Chenery, H. B. 281
Christensen, L. R. 279
Coen, R. M. 214
Crotty, J. R. 20
Denison, E. F. 220, 354
Diaz Alejandro, C. F. 299
Diewert, W. E. 260
Downs, A. 410
Eads, G. C. 50
Eisner, R. 225
Enthoven, A. C. 142
Feldstein, M. 182
Fishlow, A. 102
Fraumeni, B. M. 326
Frenkel, J. A. 235, 374
Frey, W. H. 421
Garn, H. A. 420
Gordon, R. J. 243
Griliches, Z. 343
Hall, R. E. 157
Harberger, A. C. 331
Harcourt, G. C. 27
Haveman, R. H. 88
Havrylyshyn, O. 233
Hawley, E. W. 317
Heston, A. 96
Hickman, B. G. 214
Hirschman, A. O. 113
Holland, D. M. 320
Jorgenson, D. W. 268, 326
Katznelson, I. 117
Keating, M. O. 213
Kendrick, J. W. 232
Kenyon, P. 26
Kindleberger, C. P. 293
King, M. A. 72
Klein, B. 356
Kosters, M. 187
Krueger, A. O. 288
Kulik, J. C. 132
Lau, L. J. 268
Lazear, E. P. 203

Levine, M. E. 318
McKinnon, R. I. 382
Madden, J. F. 191
Magee, S. P. 368
Mark, J. A. 353
Matthaei, J. A. 198
Melody, W. H. 393
Meltzer, A. H. 258
Michael, R. T. 203
Mieszkowski, P. 89
Morley, S. A. 110
Morey, M. J. 35
Morris, M. D. 109
Mussa, M. L. 374
Myers, S. C. 320
Nadiri, M. I. 349
Nelson, R. R. 62
Noll, R. G. 316
Pack, J. R. 422
Palmer, R. W. 123
Panzar, J. C. 311
Piekarz, R. 68
Polinsky, A. M. 363
Pollak, R. A. 273
Quinlan, D. C. 209
Rao, R. K. S. 368
Rasmussen, D. W. 415
Rees, A. 340
Reich, M. 128
Reinhardt, U. E. 149
Reynolds, L. G. 91
Rosen, H. S. 171
Salemi, M. K. 41
Scitovsky, T. 1
Shackelford, J. A. 209
Shoven, J. B. 164, 188
Siegfried, J. J. 29
Sims, C. A. 250
Small, K. A. 410
Smythe, D. W. 403
Standish, T. K. 398
Stegman, M. A. 415
Steinberg, D. 132
Stoker, T. M. 268
Summers, R. 96
Sweeney, G. H. 29
Tarshis, L. 10
Tauchen, G. E. 41
Temin, P. 301
Terleckyj, N. E. 55
Thurow, L. C. 82
Trebing, H. M. 388
Ulen, T. S. 306
von Furstenberg, G. M. 177
Wilson, S. 132
Yellen, J. L. 15

CSWEP

The Committee on the Status of Women in the Economics Profession

established in 1971 by the American Economics Association

☐ **Provides a roster service**

- Maintains and updates a list of women economists that includes fields of specialization and professional accomplishments
- Provides information of women economists at nominal cost to employer organizations

☐ **Publishes a newsletter three times a year**

- Reports on activities of the committee
- Provides brief job listings
- Prints calls for papers and mongraphs
- Lists publications and conferences of interest to women
- Presents news of other organizations

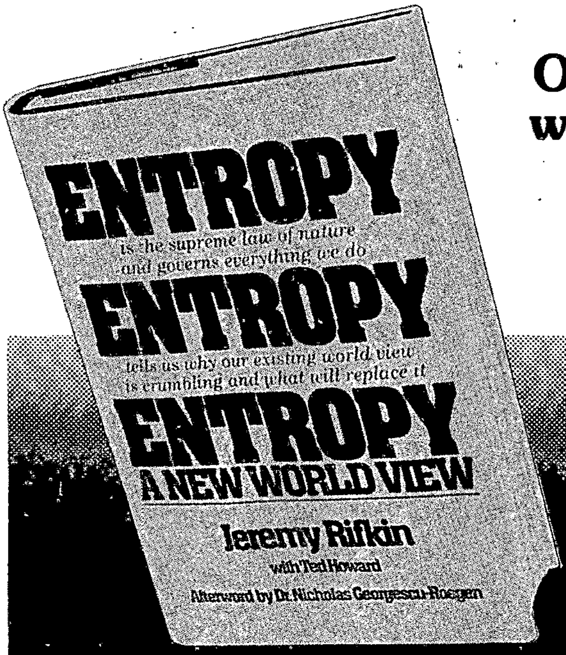
☐ **Sponsors sessions on research related to women's issues at the American and regional economics association meetings**

☐ **Collects and distributes information on the status of women in the profession**

Membership is open to women and men. To join please send $5.00 (or more) to:

Dr. Nancy Ruggles
Institute of Social and Policy Studies
Yale Station
New Haven, CT 06520
(203) 425-8255

For information about the CSWEP roster contact Dr. Ruggles at the above address.


ENTROPY
is the supreme law of nature and governs everything we do
ENTROPY
tells us why our existing world view is crumbling and what will replace it
ENTROPY
A NEW WORLD VIEW
Jeremy Rifkin
with Ted Howard
Afterword by Dr. Nicholas Georgescu-Roegen

Once in a great while a concept changes the course of history.

The Entropy Law states that all energy flows from orderly to disorderly, from usable to unusable. In his new book "ENTROPY", Jeremy Rifkin explains how this inexorable process governs every aspect of our life.

"It describes with clarity the intellectual wreckage that must be cleared away before human beings can construct a social order that is at peace with the natural order." —*Quest 80*

"It has compelled me to re-evaluate much of the safe and comfortable thinking which governs our day to day lives." —*Senator Mark Hatfield*

"A cogent and urgent arguement . . . sobering and hard hitting." —*Publisher's Weekly*

". . . most articulate and comprehensive . . . I do believe the book should be widely read and taken to heart . . ." —*Hugh Downs, ABC-News*

". . . penetrates to the heart of the energy equation in human history. Rifkin has painted a picture of the moral order that must eventuate." —*Theodore A. Wertime, Smithsonian Institution*

On Sale Now at Your Book Store. VIKING. $10.95

Over 100,000 students are learning basic economics from Edwin Dolan.... Join them!

BASIC ECONOMICS, 2nd Edition
Edwin G. Dolan, *George Mason University*

This widely adopted text features distinctive coverage of current interest topics, outstanding graphics, and the most complete package of ancillaries available: an all-new Study Guide with a unique section on careers in economics, a loose-leaf format Instructor's Manual, two Test Banks (also available on computer cards or tapes), 50 Transparency Acetates, and a Semiannual Newsletter. 1980, 736 pp., cloth

ALSO AVAILABLE IN TWO VOLUMES
Basic Macroeconomics, 2nd Edition 1980, 410 pp., paper
Basic Microeconomics, 2nd Edition 1980, 475 pp., paper

NEW IN 1981

An Introduction to Money and Banking, 4th Edition
Colin D. Campbell, *Dartmouth College,*
& Rosemary Campbell
1981, 608 pp., cloth, IM

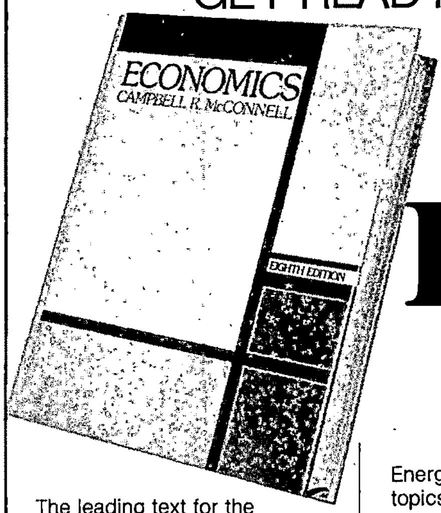
Fundamentals of Managerial Economics
James L. Pappas, *University of Wisconsin,*
& Eugene F. Brigham, *University of Florida*
1981, 600 pp., cloth, IM, SG

Macroeconomics, 2nd Edition
J. C. Poindexter, *North Carolina State University*
January 1981, 560 pp., cloth, IM

dp The Dryden Press dp
a division of Holt, Rinehart and Winston
901 N. Elm St., Hinsdale, Ill. 60521

McConnell's
Economics

GET READY FOR THE New Eighth Edition



The leading text for the principles of economics course is now available in a new, completely revised and updated eighth edition.

You'll find the new edition of McConnell to be as logically organized and well written as its predecessors, with the outstanding learning aids that have helped over four million students through their first course in economics.

A Meticulous Revision

The eighth edition reflects Professor McConnell's extensive, ongoing teaching experience, as well as his responsiveness to reviewer and user comments and suggestions. This revision features a completely new chapter on "The Economics of Energy." It covers such current topics as the tax revolt, the subterranean economy, the political business cycle, supply-side economics, and the theory of rational expectations.

Complete Support for You and Your Students

The McConnell package for the eighth edition is as complete and effective as ever:

- The *INSTRUCTOR'S MANUAL and TEST FILE* has been thoroughly redesigned to create a more open, useful format; it retains the teaching aids and test questions that have proven so successful in previous versions, and now includes 500 completely new multiple-choice questions.
- The *STUDY GUIDE* features the review and self-test aids that students have found helpful in the past, plus an invaluable new glossary.
- A unique, self-paced study guide, *ECONOMIC CONCEPTS: A Programmed Approach, 6/e* is also available.
- *TRANSPARENCY MASTERS* (newly revised) and *OVERHEAD TRANSPARENCIES* round out this unparalleled teaching and learning package.

ECONOMICS, Eighth Edition
Campbell R. McConnell, University of Nebraska-Lincoln
1981, 897 pages, $19.95 (tent.)

Price subject to change.

McGraw-Hill
COLLEGE DIVISION
McGraw-Hill Book Company
1221 Avenue of the Americas
New York, N.Y. 10020

New and forthcoming titles . . .

Urban Economics and Public Policy, *Second Edition*

JAMES HEILBRUN, *Fordham University*

Thoroughly revised and updated, **Urban Economics and Public Policy, Second Edition,** provides a lucid, nonmathematical introduction to the study of urban economics. Professor Heilbrun describes the growth of cities and metropolitan areas and analyzes major policy issues—poverty, housing, and public finance. Important features of the new edition include the addition of two chapters on transportation and expanded treatment of public finance. Running through the entire book is a connecting theme: the influence of technological innovation, rising living standards, and changing population size on the spatial organization of cities and metropolitan areas and on the major economic problems they face.

Praise for the First Edition:

"The overall content, organization and style should make this a leading publication in its field. The book's clarity together with its emphasis on public policy will have broad appeal to planners, economists, public administrators and other 'urbanologists.' "
—Andrew A. Dzurik, Florida State University.

March 1981 about 480 pages hardcover

Economic Issues Today **Alternative Approaches,** *Second Edition*

ROBERT B. CARSON, *State University of New York, College at Oneonta*

The Second Edition of Carson's thought-provoking text examines sixteen vital and current economic issues from three points of view—conservative, liberal, and radical. These arguments are presented from a first-person advocacy position to give students the impact of a debate that is free of jargon, excessive citations, or footnotes.

In the Second Edition four new issues have been added:

- The Macroeconomy Examined: Are We Depression-proof?
- Stabilization Policy: How Well Has It Been Working?
- Crisis in International Trade: Rising Deficits and Shrinking Dollars
- National Economic Planning: Is Central Control Inevitable?

The twelve issues retained from the previous edition have been revised and updated to reflect important economic events and changes in policy directions; to include more empirical information; and to incorporate a greater number of graphs and tables. **Economic Issues Today** will stimulate class discussion and student interest in your courses in Principles of Economics or Current Economic Issues.

Instructor's Manual available.

October 1980 342 pages paperbound

Also available:

Microeconomic Issues Today: Alternative Approaches
(Issues 1-8 of **Economic Issues Today,** with an Introduction and Conclusion)
182 pages paperbound $5.95

Macroeconomic Issues Today: Alternative Approaches
(Issues 9-16 of **Economic Issues Today,** with an Introduction and Conclusion)
182 pages paperbound $5.95

St. Martin's Press

Macroeconomics

ANTHONY S. CAMPAGNA, *University of Vermont*

In **Macroeconomics** Professor Campagna presents the basic analytical tools of macroeconomic theory without burdening the analysis with unnecessary mathematical detail. A major emphasis throughout the text is on the examination and evaluation of practical policy matters. Both theory and policy are placed in an historical perspective, providing students with an appreciation of how they interact. In addition, the text surveys the criticisms of and alternatives to the prevailing orthodoxy—the whole spectrum of economic thought from Friedman and Galbraith to the radicals and post-Keynesians. By taking this undogmatic approach, the book encourages an awareness of the evolutionary nature of macroeconomics and of the controversies that surround it today.

Instructor's Manual available.
Study Guide available.

February 1981 about 600 pages hardcover

Introductory Statistics for Business and Economics

RICHARD K. MILLER, *Washington and Jefferson College*

Incorporating a wealth of examples from the real world, **Introductory Statistics for Business and Economics** consistently presents statistical methods as vehicles for making business and economic policy decisions. The text assumes only high-school algebra as a prerequisite and explains the logic underlying statistical techniques with unusual clarity.

The coverage of the text is extensive enough to meet the needs of any introductory course. Its flexibility of organization allows instructors to focus the course as they choose: the basic statistical concepts—descriptive statistics, probability, and statistical inference—are treated in the first eleven chapters; the remaining nine chapters offer a variety of topics from which to choose and may be covered in any desired sequence.

In addition to the numerous worked-out examples throughout the text, Professor Miller provides approximately 500 end-of-chapter problems. An appendix includes necessary statistical tables and answers to selected problems.

Instructor's Manual available.
Study Guide available.

November 1980 445 pages graphs, charts, tables hardcover

To request a complimentary examination copy for adoption consideration, please write (specifying course, present text, and approximate enrollment) to:

St. Martin's Press ■ P.O. Box 5352 ■ New York, N.Y. 10163

Liberty*Press*
Liberty*Classics*

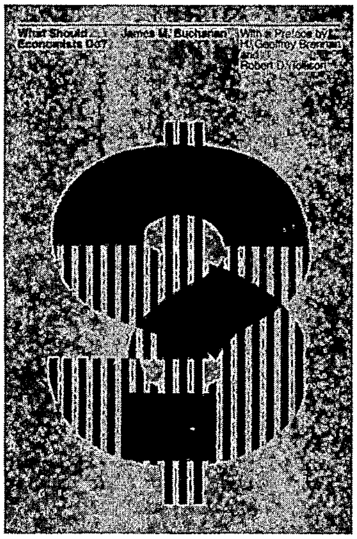
What Should Economists Do?

By James M. Buchanan

This collection brings together Buchanan's important essays on method, many of them previously unpublished. Such a volume, note H. Geoffrey Brennan and Robert D. Tollison in their preface, "provides relatively easy access to a group of significant papers on methodology in economics, written by a man whose work has spawned a methodological revolution in the way economists and other scholars think about government and governmental activity."

James M. Buchanan is University Distinguished Professor and General Director of the Center for Study of Public Choice, Virginia Polytechnic Institute and State University. As a founder of the burgeoning subdiscipline of public choice, and as a public finance theorist and father of a modern school of public finance, Buchanan's work has had worldwide recognition. Hardcover $8.00, Paperback $3.50.

We pay postage, but require prepayment, on orders from individuals. Please allow four to six weeks for delivery. To order this book, or for a copy of our catalog, write:
Liberty*Press*/Liberty*Classics*
7440 North Shadeland, Dept. 651
Indianapolis, Indiana 46250


What Should Economists Do?
James M. Buchanan
With a Preface by H. Geoffrey Brennan and Robert D. Tollison

Now Available —————— Conference Series No. 21

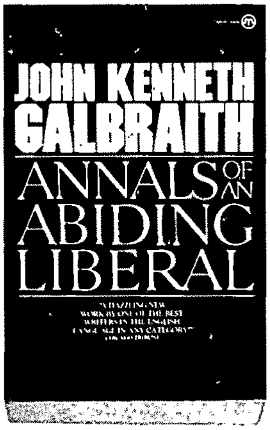
The Regulation of Financial Institutions

Barth
Brandt
DeMagistris
Durkin
Hammond
Maisel
Peterson
Shapiro
Tucker
Benston
Cordes
Dunkelberg
Yezer
Greenwald
Kopcke
Merton
Schafer
Shay
Woglom

Proceedings of a conference held in October 1979 sponsored by the Federal Reserve Bank of Boston and the National Science Foundation.

Available on request without charge from:
The Federal Reserve Bank of Boston
600 Atlantic Avenue, Boston, MA 02106
Research Department

"A dazzling new work by one of the best writers in the English language in any category!"
—*Chicago Tribune*



Wit and animation, erudition and penetrating insights are the unmistakable hallmarks of Galbraith's writing—and never more so than in this spirited collection which probes the current preoccupations of the American community. Whether discussing the recent economic climate or contemporary swindlers, Nixon's memoirs or Evelyn Waugh's novels, Galbraith's pithy and pointed essays will challenge and charm readers of all political persuasions. "Galbraith has never been more diverse in subject matter... never more Galbraithian in pure, enjoyable artistry!"
—*Boston Globe*

MERIDIAN / F544 / $5.95

At all bookstores or send $5.95 plus $.75 for postage and handling to: New American Library, P.O. Box 999, Bergenfield, N.J. 07621

Write to the NAL Education Department at the following address for a free History and Social Studies catalog.

NEW AMERICAN LIBRARY NAL TIMES MIRROR 1633 Broadway New York 10019

New perspectives on ECONOMICS from Winthrop

INVESTMENT ANALYSIS AND MANAGEMENT

Stanley S.C. Huang

Rider College

Excellent for undergraduate investments courses, this new text features detailed material on investment management, security analysis, and the theoretical and practical aspects of modern portfolio management.

1981/casebound/576pp/$19.95/Solutions Manual

PUBLIC SECTOR ECONOMICS

Robin W. Boadway

Queen's University, Kingston, Ontario

For courses on Public Finance—
"This book combines the conventional microeconomic approach to public finance with some of the recent theoretical developments in the field. The author uses both normative and positive analysis. . ."

Economic Forum

1979/casebound/480pp/$19.95

INTERNATIONAL ECONOMICS

Robert J. Carbaugh

University of Wisconsin, Eau Claire

Designed for students with an understanding of basic economics, this introductory text skillfully combines theory-policy-institutional material in a real-world context. Emphasis is on current policies such as nontariff trade barriers, the OPEC oil cartel, multinational corporations, and exchange rate systems.

1980/casebound/448pp/$15.95

LABOR ECONOMICS: THE EMERGING SYNTHESIS

Robert M. Fearn

North Carolina State University

Fearn's mainstream text is a mjaor new choice for a first course in labor economics. Utilizing a topical approach to stimulate student interest, Fearn develops the theory necessary for each topic before proceeding to more complex issues, such as welfare, work incentives, migration, minimum wage, and the effects of labor unions and women's liberation.

1981/casebound/approx.304pp/$18.95

COMPARATIVE ECONOMICS: NATIONAL PRIORITIES, POLICIES, AND PERFORMANCE

Jahangir Amuzegar

Executive Director, International Monetary Fund

Current factual material, a unique multidisciplinary approach, and a lucid presentation distinguish this new text for undergraduate courses. The text covers decision-making processes in market oriented industrial countries, centrally directed economies, and developing nations.

1981/casebound/480pp/$17.95

For further information, or to order examination copies, please write to Ms. Sara Black, Dept. 4680.

WINTHROP PUBLISHERS, INC.

17 Dunster Street • Cambridge, MA 02138

Wealth, Income, and Inequality

Second Edition

Edited by A. B. ATKINSON, Professor of Economics, London School of Economics. This extensively revised edition incorporates a broad range of material, including a large amount of recent research, such as the work of the Royal Commission on the Distribution of Income and Wealth. Principal topics discussed are the definition of basic concepts, the measurement of the distribution of income and wealth, and analysis of the factors influencing that distribution.

December 1980 450 pp. cloth $49.95 paper $24.00

The Political Economy of the Third Sector

Co-operation and Participation

Edited by ALASDAIR CLAYRE. Essays in this book focus on the possible development of a "third sector" of industry, distinct from private industry and from public nationalized enterprise, in which management is by the workers. Contributors discuss arguments for and against participation in such a plan, the degrees of legislation necessary, the problems of raising capital, and the effect of such a sector on industrial relations.

1980 200 pp. cloth $29.50 paper $14.95

Seeds of Plenty, Seeds of Want

A Critical Analysis of the Green Revolution

ANDREW PEARSE, Co-Director, Participation Programme of United Nations Research Institute for Social Development. The Green Revolution was an international campaign aimed at increasing the productivity of land by means of the introduction of a science-based technology in the production of foodgrains. Pearse discusses the failure of the project and indicates that development strategies have not effectively incorporated economic aspects of rural population. He suggests, further, new policies and strategies whereby productivity can be increased without gross polarization of wealth.

1980 262 pp. $22.50

Economic Sanctions and International Enforcement

Second Edition

MARGARET P. DOXEY, Professor of International Relations, Trent University. From a review of the first edition: "An excellent study. A lucid survey of cases of imposition of sanction...is followed by a judicious and thorough analysis of the factors bearing upon the effectiveness of sanctions.... This book deserves to be widely read."—*International Affairs.* For this new edition Professor Doxey takes into account new cases of economic sanctions since the first edition was published in 1970, including the changes in the situation in Southern Africa and the Arab oil embargo of 1973.

1980 160 pp. $22.00

Inequality and Economic Development in Malaysia

DONALD R. SNODGRASS, Harvard Institute of International Development. The problem of inequality and economic development is one which Malaysia shares with many developing countries. In this study, Professor Snodgrass focuses on the ethnic dimensions of the problem, details economic policy since 1969, and appraises the efforts of the Malaysian Government to deal with economic inequality. Chapters are devoted to fiscal redistribution, rural and agricultural development policies, urban and industrial development policies, education, and general policies.

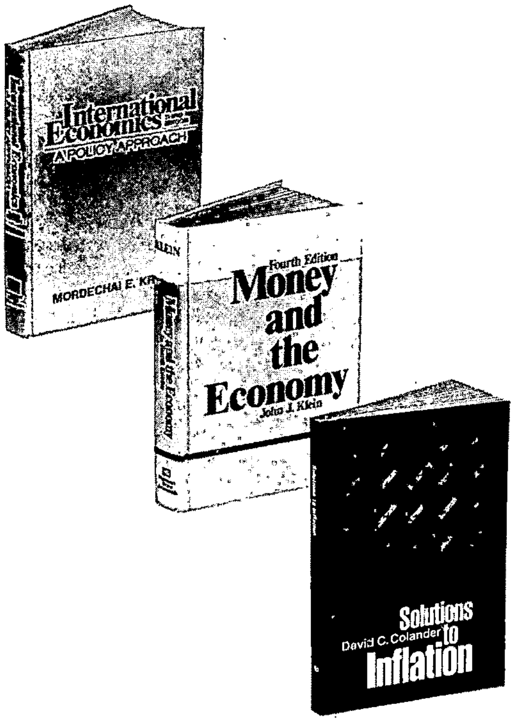
1980 340 pp.; numerous tables cloth $32.50 paper $17.95

Prices and publication dates are subject to change.

OXFORD UNIVERSITY PRESS
200 Madison Avenue, New York, N.Y. 10016

SINCE 1478

The latest in economic



Money and the Economy

FOURTH EDITION

JOHN J. KLEIN, Georgia State University

The new Fourth Edition of this highly respected textbook continues to offer a comprehensive explanation and analysis of financial institutions and markets, theory, and economic policy. Professor Klein's aim is to give students not a financier's view of the system, or the Federal Reserve's view, or the bias of any school of economic theory, but rather a well-integrated, overall perception of the financial system and the theory on which it is based.

598 pages

Solutions to Inflation

DAVID C. COLANDER, University of Miami

A collection of 29 stimulating articles, this reader focuses on current economic policy options—or solutions—for dealing with inflation. Among the topics discussed are indexing, wage-price guidelines and controls, establishing a "free market" in money, and incentive-based incomes policies. The articles—a number of which are published here for the first time—include selections by such leading economists as Friedman, Galbraith, Hayek, Solow, Okun, and Weintraub.

Paperbound. 224 pages

International Economics
A Policy Approach

THIRD EDITION

MORDECHAI E. KREININ,
Michigan State University

In this revised and updated edition, Professor Kreinin presents a clear and comprehensive introduction to international finance and trade. As in the earlier editions, there is a firm emphasis on policy and analysis of general principles, focused directly on the international scene. Domestic problems are discussed only insofar as they are relevant to their international counterparts, but the author uses the United States—the most important trading nation in the world—as the reference point for evaluating policy alternatives.

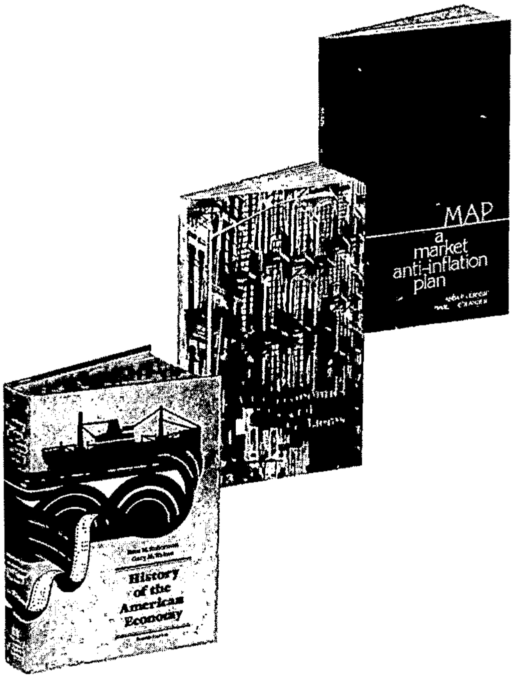
464 pages

HBJ

HARCOURT BRACE JOVANOVICH, INC.

New York • Chicago • San Diego • Atlanta • San Francisco

incentives. . .from HBJ



History of the American Economy

FOURTH EDITION

ROSS M. ROBERTSON,
Late of Indiana University
GARY M. WALTON,
Washington State University

This textbook offers a highly readable account of the major events, institutions, ideas, and creative works that have shaped our economic system from colonial times to the present. In preparing the Fourth Edition, Professor Walton reconsidered the entire text, revising, reorganizing, and rewriting where necessary to make the book more balanced and accessible. A consistent emphasis on principles, an engaging style, and judicious selection of materials help clarify students' understanding of major economic developments in American history.
586 pages

MAP

A Market Anti-Inflation Plan

ABBA P. LERNER, Florida State University
DAVID C. COLANDER, University of Miami

MAP offers a provocative answer to a persistent problem: inflation/stagflation. Synthesizing many divergent trains of thought, Professors Lerner and Colander suggest a new way of looking at inflation—in fact, at all macroeconomics. Their plan involves stabilizing average prices while allowing all actual prices and wages to adjust to changing tastes, technologies, and resources so that markets are cleared. In their clear, careful explanation of the plan—and of the policy adjustments it would involve—the authors provide a major economic statement for the 1980s and beyond.
Paperbound. 128 pages

The Economics of Social Problems

JULIAN LE GRAND,
London School of Economics
RAY ROBINSON, University of Sussex

This book presents key principles of economics in a relevant social context: the authors develop basic concepts and relevant theory simultaneously and apply them to such current social issues as health, education, housing, crime, congestion, and the energy crisis. Consistent use of a rigorous theoretical framework helps students acquire essential skills and a sound understanding of the generality of economic analysis.
Paperbound. 200 pages

HARCOURT BRACE JOVANOVICH, INC.

New York • Chicago • San Diego • Atlanta • San Francisco

Economic Growth and Resources

Proceedings of the Fifth World Congress of the International Economic Association Held in Tokyo, 1977

Volume One of *Economic Growth and Resources* provides the plenary session addresses of the IEA's Fifth World Congress, attended by some fifteen hundred participants. The remaining four volumes present papers from the specialized groups.

Volume One: The Major Issues
edited by Edmond Malinvaud
1979 ISBN 0-312-23314-0 278 pp., cloth $37.50

Volume Two: Trends and Factors
edited by R.C.O. Matthews
1980 ISBN 0-312-23315-9 335 pp., cloth $37.50

Volume Three: Natural Resources
edited by Christopher Bliss
1980 ISBN 0-312-23316-7 210 pp., cloth $37.50

Volume Four: National and International Economic Policies
edited by Irma Adelman
1979 ISBN 0-312-23317-5 384 pp., cloth $37.50

Volume Five: Problems Related to Japan
edited by Shigeto Tsuru
1979 ISBN 0-312-23318-3 412 pp., cloth $37.50

To order these books for your library
at 20% discount, please write:

St. Martin's press
Post Office Box 5352 New York, NY 10163

PRAEGER

entering our fourth decade

ENERGY FUTURES OF DEVELOPING COUNTRIES
The Neglected Victims of the Energy Crisis
edited by **Harlan Cleveland**
ca. 128 pp., 1980, $18.95, ISBN 0-03-058669-0

PROGRESS AND SURVIVAL
An Essay on the Future of Mankind
Emile Benoit, edited by **Jack Benoit Gohn**
144 pp., 1980, $16.95, ISBN 0-03-056911-7

FOREIGN INVESTMENT CODES AND THE LOCATION OF DIRECT INVESTMENT
Robert E. Grosse
174 pp., 1980, $19.95, ISBN 0-03-057024-7

THE ETHICAL BASIS OF ECONOMIC FREEDOM
edited by **Ivan Hill**
352 pp., 1980, $9.95, ISBN 0-03-057732-2, Paperback

THE ECONOMY OF SAUDI ARABIA
Donald M. Mollver and **Paul J. Abbondante**
ca. 200 pp., 1980, ca. $19.95, ISBN 0-03-057004-2

THE INDONESIAN ECONOMY
edited by **Gustav F. Papanek**
ca. 300 pp., 1980, ca. $21.95, ISBN 0-03-057429-3

TROUBLE IN EDEN
A Comparison of the British and Swedish Economies
Eli Schwartz
160 pp., 1980, $18.95, ISBN 0-03-057032-8

THE ITALIAN ECONOMY
Donald C. Templeman
ca. 400 pp., 1980, ca. $22.95, ISBN 0-03-057612-1

Send for our FREE catalogue!
Order from:
PRAEGER PUBLISHERS
521 Fifth Avenue
New York, New York 10175

Of Economic Importance...

New 1981 Titles From P-H

THE ECONOMIC PROBLEM, Sixth Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 672 pp. (est.) Cloth

UNDERSTANDING MACROECONOMICS, Seventh Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 384 pp. (est.) Paper

UNDERSTANDING MICROECONOMICS, Fifth Edition
Robert L. Heilbroner, New School for Social Research
Lester C. Thurow, Massachusetts Institute of Technology
1981 352 pp. (est.) Paper

MONEY, BANKING AND ECONOMIC ANALYSIS, Second Edition
Thomas D. Simpson, George Washington University
1981 502 pp. (est.) Cloth

ECONOMICS OF THE FIRM: Theory and Practice, Third Edition
Arthur A. Thompson, Jr., Univ. of Alabama
1981 656 pp. (est.) Cloth

MARKETS, DECISIONS, AND ORGANIZATIONS: Intermediate Microeconomic Theory
Roger A. McCain, Temple University
1981 620 pp. (est.) Cloth

BUSINESS, GOVERNMENT, AND THE PUBLIC, Second Edition
Murray L. Weidenbaum, Washington University
1981 336 pp. (est.) Cloth

INTERMEDIATE MACROECONOMICS
William A. Kelly, Jr., Pennsylvania State University
1981 500 pp. (est.) Cloth

BENEFIT-COST ANALYSIS OF GOVERNMENT PROGRAMS
Edward M. Gramlich, Univ. of Michigan
1981 352 pp. (est.) Cloth

MACROECONOMICS: Theory, Evidence and Policy, Second Edition
Frank C. Wykoff, Pomona College
1981 550 pp. (est.) Cloth

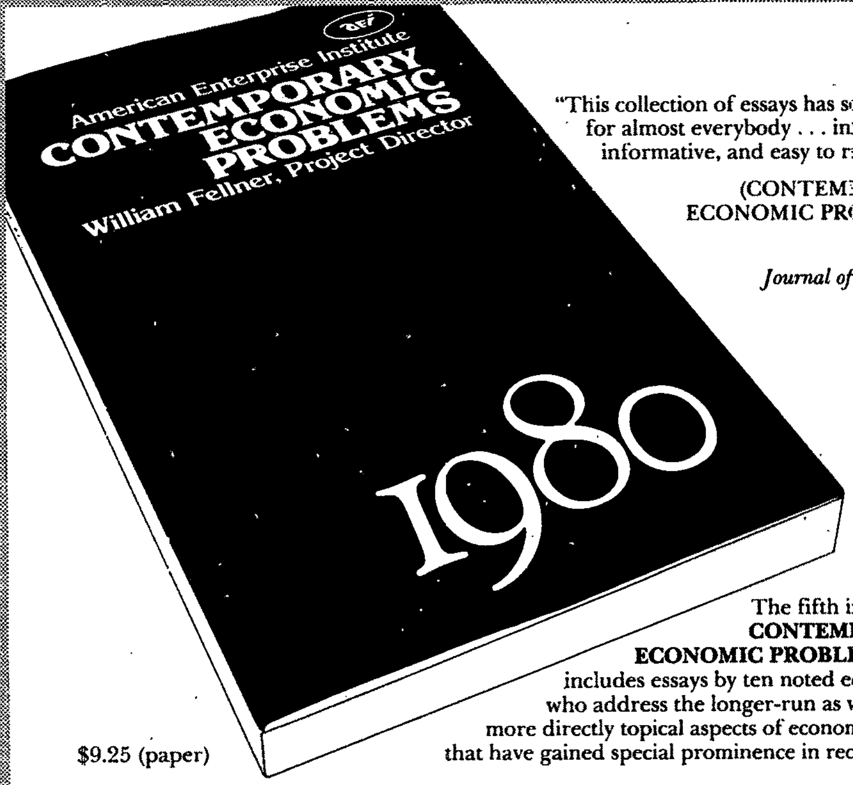
COPING WITH CAPITALISM: The Economic Transformation of the United States, 1776-1980
Roger L. Ransom, University of California — Riverside
1981 256 pp. (est.) Paper

MACROECONOMICS FOR MANAGEMENT
Mary Louise Hatten, Boston College
1981 384 pp. (est.) Cloth

REFORMING REGULATION: Processes and Problems
Lawrence J. White, New York University
1981 128 pp. (est.) Cloth

For further information, or to order or reserve examination copies, please write: Robert Jordan, Dept. J-680, Prentice-Hall, Inc., Englewood Cliffs, NJ 07632
For "SUPER-QUICK" Service, dial TOLL FREE (800) 526-0485,* between 9:00 a.m.-4:30 p.m. EST, Mon.-Fri.
*not applicable in New Jersey, Alaska, Hawaii, or Puerto Rico

Prentice-Hall


AEI
American Enterprise Institute
CONTEMPORARY ECONOMIC PROBLEMS
William Fellner, Project Director
1980

"This collection of essays has something for almost everybody . . . interesting, informative, and easy to read . . . ."

(CONTEMPORARY
ECONOMIC PROBLEMS
1979)

Journal of Economic Literature

The fifth in a series, **CONTEMPORARY ECONOMIC PROBLEMS 1980** includes essays by ten noted economists who address the longer-run as well as the more directly topical aspects of economics issues that have gained special prominence in recent years.

$9.25 (paper)

WILLIAM FELLNER
Demand Disinflation: What If Gradualism Should Fail Despite Its Merits?

PHILLIP CAGAN
The New Monetary Policy and Inflation

HERBERT STEIN
Achieving Credibility

WILLIAM FELLNER
Corporate Asset-Liability Decisions in View of Low Market Valuation of Equity

RUDOLPH G. PENNER
The Future Growth of Government Budgets

GOTTFRIED HABERLER
The Dollar in the World Economy: Recent Developments in Perspective

GEOFFREY H. MOORE
Inflation and Statistics

JACK A. MEYER
Wage and Benefit Trends Under the Carter Administration Guidelines

MARK PERLMAN
Prices, Technological Change, and Productivity in the American Health Care Industry

BARRY R. CHISWICK
Review of "Counting the Labor Force," the Report of the National Commission on Employment and Unemployment Statistics

D. GALE JOHNSON
The World Food Situation: Developments During the 1970's and Prospects for the 1980's

TO ORDER: Send your check made out to "American Enterprise Institute" and your name and address to:

American Enterprise Institute
1150 Seventeenth Street, N.W.
Washington, D.C. 20036

or call AEI's toll-free number: (800) 424-2873

CHICAGO

Since 1891, Publishers of Scholarly Books and Journals

Asset Accumulation and Economic Activity

Reflections on Contemporary Macroeconomic Theory

James Tobin

"*Asset Accumulation and Economic Activity* is a cogent statement of the views of one of the most respected and thoughtful macroeconomists in the field, and will serve as a convenient reference and focal point from which a base future discussions."—Stanley Fischer, MIT

Cloth 112 pages $13.00

Essays in Applied Price Theory by Reuben A. Kessel

Edited by R. H. Coase and Merton L. Miller

"The seminal articles represented in this collection have exerted a lasting influence on our understanding of economic markets—from medicine to milk to mortgages and beyond."—Milton Friedman

Cloth 384 pages $20.00

Problems in Price Theory

David de Meza and Michael Osborne

Students often find microeconomic theory the most difficult part of economics. This set of problems and solutions is designed to enable students to become proficient by working out problems on an ongoing basis.

Cloth 320 pages $25.00

The System-Wide Approach to to Microeconomics

Henri Theil

Theil constructs an overview of recent theory and research with empirical illustrations, to show the increasing parallelism between the theories of the consumer and the firm.

Cloth 288 pages $23.00

The Road to Serfdom

Friedrich A. Hayek

Nobel Laureate Friedrich Hayek's classic warning to the free nations of the world that the extended collectivism implicity in "Social planning" is incompatible with the operation of democratic institutions.

Cloth 256 pages $12.50
Paper $4.95

Human Capital

A Theoretical and Empirical Analysis, with Special Reference to Education

Second Edition

Gary S. Becker

"Gary Becker's *Human Capital* is without question the most influential volume in the large literature on labor earnings."—Robert T. Michael, Department of Economics, Stanford University

Paper 288 pages $7.50

Competition and Coercion

Blacks in the American Economy, 1865–1914

With a new Preface

Robert Higgs

Higgs show that competition as well as coercion characterized the post-emancipation period, and that blacks made substantial economic progress in spite of political and social obstacles.

Paper 224 pages $6.95

THE UNIVERSITY OF CHICAGO PRESS

Chicago 60637

Quality texts that bring the real world of contemporary economics into the heart of the classroom...

Coming Soon!

Economics: Principles in Action, Third Edition
by Philip C. Starr, Chaffey College

Captivatingly written . . . flexible . . . thoroughly updated, yet still concise. Tax revolt, supply side fiscalism, and post-Keynesian economics are just some of the ideas, policy questions, and issues that have been blended into this extensively fine-tuned third edition of Starr's highly readable text. Designed for the one-term or two-quarter economics survey course and ideal for students with little or no formal economics education.

February 1981. Approximately 424 pages. 7-3/8 x 9-1/4. Paperbound. Student Supplement. Instructor's Manual.

Applications of Macroeconomics and Applications of Microeconomics
both volumes by Robert Paul Thomas, University of Washington

A well-devised applications book to complement your principles of economics text (one volume for macroeconomics . . . one volume for microeconomics). Thomas' new supplementary paperbacks apply the important concepts of economics (one from the macro point of view, the other from the micro point of view) with more than 30 relevant economic and social applications. Each application is accompanied by a preview, key points, and suggested readings.

Both volumes: February 1981. Approximately 260 pages each. 6-1/2 x 9-1/4. Paperbound. Instructor's Manual.

Information!

For more information on these texts or for a copy of the Wadsworth Economics catalogue, write: **Box AER-W-DEC80,**

Wadsworth Publishing Company
Ten Davis Drive
Belmont, California 94002

Now Available!

Economics: Concepts, Themes, Applications
by Don R. Leet and John A. Shaw, both California State University, Fresno

The one-term intro text that offers a balanced coverage of macro and micro concepts that's reinforced with vivid and timely applications, examples and issues, such as inflation, Disneyland, property rights, and OPEC.

450 pages. 7-3/8 x 9-1/4. Paperbound. Student Workbook. Instructor's Manual.

The Making of Economics, Second Edition
by E. Ray Canterbery, Florida State University

The provocative supplementary text that's become a minor classic in the History of Economic Thought.

340 pages. 6-1/2 x 9-1/4. Paperbound.

Intermediate Microeconomics
by Steven T. Call, Metropolitan State College; and William L. Holahan, University of Wisconsin, Milwaukee

A skillful blend of analysis and applications that provides intermediate students with a rich and interesting look at microeconomics.

514 pages. 7-3/8 x 9-1/4. Casebound. Instructor's Manual.

Contemporary Labor Economics and Labor Relations, Second Edition
by Juanita Kreps, former Secretary of Commerce, Washington, D.C.; Philip L. Martin, University of California, Davis; Richard Perlman, University of Wisconsin, Milwaukee; and Gerald G. Somers, formerly of University of Wisconsin, Madison

The labor economics and labor relations text that's as current as your students' questions . . . thoroughly revised, updated, and reorganized.

484 pages. 7-3/8 x 9-1/4. Casebound. Instructor's Manual.

Urban Public Economics
by William B. Neenan, University of Michigan

The urban economics text that emphasizes the public policy aspect. (Duxbury Press)

Approximately 400 pages. 6-3/8 x 9-1/4. Casebound. Available Spring 1981.

WP Wadsworth Publishing Company

The Fraser Institute

Timely new titles on
Canada's important
public policy issues

PRIVATIZATION
Theory and Practice
256 pages, $12.95 ISBN 0-88975-036-X

RENT CONTROL
Myths & Realities
368 pages, $7.95 ISBN 0-88975-033-5

UNIONS AND THE PUBLIC INTEREST
Collective Bargaining in the Government Sector
128 pages, $5.95 ISBN 0-88975-022-X

gage PUBLISHING LIMITED

IS THE EXCLUSIVE NATIONAL DISTRIBUTOR
FOR ALL FRASER INSTITUTE TITLES.

PUBLISHED BY THE FRASER INSTITUTE, 626 BUTE STREET, VANCOUVER, BRITISH COLUMBIA, CANADA V6E 3M1. AVAILABLE AT BOOKSTORES OR BY MAIL FROM THE INSTITUTE.
Please write to the Publications Manager if you would like to receive news of our new and forthcoming titles. Libraries and other institutions are invited to place standing orders for Fraser Institute books.

Harper & Row's all-new texts...

GOLDFELD & CHANDLER
The Economics of Money and Banking ***Eighth Edition***
Instructor's Manual.
February 1981. 664 pages (tentative). ISBN 0-06-041236-4.

BRANSON & LITVACK
Macroeconomics ***Second Edition***
Instructor's Manual
March 1981. 480 pages (tentative). ISBN 0-06-040937-1

WILSON
Microeconomics
Concepts and Applications
February 1981. 416 pages (tentative). ISBN 0-912-21213-6.

SCHEUCH
Labor in the American Economy
Labor Problems and Union-Management Relations
February 1981. 576 pages (tentative). ISBN 0-06-045774-0.

LEVITAN, MANGUM, & MARSHALL
Human Resources and Labor Markets
Employment and Training in the American Economy
Third Edition
April 1981. 672 pages (tentative). ISBN 0-06-044074-0.

For examination copies, write to Suite 3D, 10 East 53d Street, New York, N.Y. 10022. Please include your course title, enrollment, and present text.

meeting the challenges of the 80's.

PETERSEN

Business and Government

Instructor's Manual.
February 1981. 480 pages (tentative). ISBN 0-06-045142-4.

GREGORY & STUART

Soviet Economic Structure and Performance ***Second Edition***

February 1981. 464 pages (tentative). ISBN 0-06-042508-3.

KELEJIAN & OATES

Introduction to Econometrics

Principles and Applications, *Second Edition*

January 1981. 384 pages (tentative). ISBN 0-06-043618-2.

HUNT & SHERMAN

Economics

An Introduction to Traditional and Radical Views
Fourth Edition

Instructor's Manual
February 1981. 672 pages (tentative). Paper. ISBN 0-06-043008-7.

HUNT

Property and Prophets

The Evolution of Economic Institutions and Ideologies
Fourth Edition

Feburary 1981. 192 pages (tentative). Paper. ISBN 0-06-043033-8.

Harper & Row

Five Economists Are Offered $200,000 Toward Law Degrees At Emory University

If you have the Ph.D. in Economics, or if you have *almost* completed all of the requirements for it, you may qualify for a John M. Olin Fellowship in Law and Economics at Emory University. Applications for the class entering in August, 1981, are now being accepted.

This three-year program, formerly conducted at the University of Miami, is designed for highly-motivated individuals interested in applying economic analysis to legal issues. It leads to the Doctor of Law Degree. Fellows may complete the program in two and one-half years by attending summer sessions.

The fellowships are valued at approximately $40,000 each. Each fellowship covers full tuition and fees and provides a stipend of at least $8,000 for Fellows who possess the Ph.D. Write:

John M. Olin Fellowship Program
Law and Economics Center
Emory University School of Law
Atlanta, GA 30322

STAFF PAPERS

Published by the International Monetary Fund in March, June, September, and December. Summaries of each paper in French and Spanish, as well as in English.

Vol. 27 September 1980 No. 3

Implications of Credit Policies for Output and the Balance of Payments .... PETER M. KELLER

Multipurpose Banking: Its Nature, Scope, and Relevance for Less Developed Countries .... DEENA R. KHATKHATE AND KLAUS-WALTER RIECHEL

Mineral Taxation Policies in Developing Countries: An Application of Resource Rent Tax .... KEITH E. PALMER

Demand for Money in Middle Eastern Countries
ANDREW D. CROCKETT AND OWEN J. EVANS

Deposit Insurance: Theory and Practice ....IAN S. MCCARTHY

The Fund Agreement in the Courts—XV ....JOSEPH GOLD

Subscription: US$9.00 a volume; US$3.00 a single issue.
Special rate to university libraries, faculty, and students: US$4.00 a volume;
US$1.00 a single issue.

Advice on payment in currencies other than U.S. dollars will be given upon request.

**The Secretary, Attention: Publications Section, International Monetary Fund,
Washington, D.C. 20431, U.S.A.**

Monetary Integration and Theory of Optimum Currency Areas in Africa

by Samuel Nana-Sinkam
1978, xii+313 pp., Clothbound, DM78/US$48.75
ISBN 90 279 7665 1

The recent surge of interest in optimum currency areas has been answered by this volume: an outstanding contribution to the field of international monetary cooperation, it presents theory and data, along with a model for solving the "which-country-which-optimum-currency-area" problem, in the context of the distinctive political and regional factors and the financial and monetary infrastructure of Africa.

Also by Samuel Nana-Sinkam:

Pays candidats au processus de développement:
Capacité d'absorption, assistance extérieur et modèles de croissance

1975, xvi+582 pp., Clothbound, DM118/US$74.00
ISBN 90 279 7665 1

Orders may be addressed to:
For USA and Canada: Mouton Publishers, Division of Walter de Gruyter, Inc., 200 Saw Mill River Road, Hawthorne, New York 10532
For all other countries: Walter de Gruyter & Co., P.O. Box 110240, D-1000 Berlin-11

mouton publishers
The Hague · Paris · New York

Front page news for

Introducing the first principles text with a supply-side emphasis—

INTRODUCTORY ECONOMICS

MICHAEL VESETH, University of Puget Sound

January, 1981 800 pages

A thoroughly modern, remarkably lucid *incentive approach* to the introductory course. Veseth brings supply-side economics out of the newspapers and into your classroom. He concentrates on the nature of economic problems and the trade-offs inherent in any solution to these problems, stressing policy rather than theory or institutions. In-depth coverage is given to international economics.

THE PACKAGE: Introductory Macroeconomics (paper). Introductory Microeconomics (paper). Coursebook (combined reader/study guide). Test File (book and computer tape formats). Instructor's Manual. Transparency Masters.

ECONOMICS

Private and Public Choice

Second Edition

JAMES D. GWARTNEY, Florida State University
and RICHARD STROUP, Montana State University,
with the assistance of A. H. STUDENMUND, Occidental College

1980 835 pages

"Excellent book, perhaps the best available. You have kept the features that led me to adopt the first edition—style, organization, and topics—particularly the unique coverage of the public choice literature as an extension of the 'economic way of thinking.' The coursebook is the best available. I particularly like the feature: "What's wrong with this way of thinking?"—a good way to stimulate classroom discussion."—Professor Robert B. Harris, University of Connecticut, Stamford

THE PACKAGE: Macroeconomics, 2e (paper). Microeconomics, 2e (paper). Coursebook (combined reader/study guide). Test Bank (book and computer tape formats). Instructor's Manual. Transparency Masters.

Address requests for examination copies to: College Department,
Academic Press, Inc., 111 Fifth Avenue, New York, N.Y. 10003.
Please state course title, enrollment, and present text.

the classroom...

MONEY, BANKING, AND ECONOMIC ACTIVITY

GAIL E. MAKINEN, U.S. General Accounting Office and Georgetown University

February, 1981 *556 pages*

An up-to-date look at money and banking, incorporating the 1980 finance laws.

MONEY AND INFLATION: A Monetarist Approach

J. HUSTON McCULLOCH, The Ohio State University

1975 *121 pages* *paper*

An unabashedly monetarist examination of the theoretical and practical aspects of monetary theory—an excellent supplement for money and banking, intermediate macroeconomics, and principles courses.

INTERMEDIATE MICROECONOMICS WITH APPLICATIONS

AROOP K. MAHANTY, University of Northern Colorado

1980 *514 pages* *Instructor's Manual*

The immediate uses of microtheory are explored through plentiful applications based on actual business practice.

ENERGY ECONOMICS AND POLICY

JAMES M. GRIFFIN & HENRY B. STEELE, University of Houston

1980 *370 pages*

The only text to explore the impact of government policy on the supply and demand for energy, particularly petroleum.

FORECASTING IN BUSINESS AND ECONOMICS

C. W. J. GRANGER, University of California, San Diego

1980 *226 pages*

A junior-senior level treatment by the recognized authority on applying econometric theory to forecasting business cycles.

ACADEMIC PRESS, INC.
A Subsidiary of Harcourt Brace Jovanovich, Publishers
NEW YORK • LONDON • TORONTO • SYDNEY • SAN FRANCISCO

Auburn House Publishing Company
Quality Professional Books

THE ADVERTISING CONTROVERSY
Evidence on Advertising's Economic Effects

Mark S. Albion, Harvard University and **Paul W. Farris,** University of Virginia with a foreword by **Stephen A. Greyser,** Harvard University

The effects of advertising on prices has been a long-debated issue between proponents of two disparate views about how advertising works. Does advertising tend to depress prices, or does it tend to increase prices? Advertising's effects on prices is one of several economic issues that the authors cover by surveying the research underlying each group's economic propositions. The effects of advertising on product innovation, demand, industry profits, and industrial concentration are also examined. They conclude that economic models usually fail to recognize the importance of the distribution sector. By including marketing considerations, Albion and Farris present a theoretical framework that incorporates the role of the retailer in an analysis of advertising's effects. As a result, the authors hope to have opened an avenue for more productive future research in this area. **1980/Hardbound/$19.95**

NON-PROFIT HOSPITALS
Their Structure, Human Resources, and Economic Importance

Thomas A. Barocci, Alfred P. Sloan School of Management, Massachusetts Institute of Technology

Non-profit hospitals, central to the delivery of health services in the United States, constitute an industry of major importance to the national economy. Rising concern over prices among the public, policy makers, and within the industry itself makes careful scrutiny of the costs and complications facing hospitals particularly timely.

Using an effective combination of case studies and macro data sets on the industry's employment and income, Professor Barocci identifies and examines the major challenges. Particular attention is paid to the factors and effects of competition, financial and regulatory restraints, and labor force size and composition. Finally, he offers a number of policy and research recommendations. **1980/Hardbound/$19.95**

AIRCRAFT INDUSTRY DYNAMICS
An Analysis of Competition, Capital, and Labor

Barry Bluestone, Peter Jordan, and Mark L. Sullivan, The Social Welfare Research Institute, Boston College

Utilizing information from numerous personal interviews and empirical data from a multitude of sources, the authors analyze the factors that influence capital investment behavior and employment in the volatile aircraft and parts industry. Market structure and the nature of competition, technology, and government and trade union relations are examined in order to determine how each affects investment decision-making and capital formation. The results of this examination also yield insights into the industry's prospects for the 1980s. **1980/Hardbound/$19.95**

THE RETAIL REVOLUTION
Market Transformation, Investment, and Labor in the Modern Department Store

Barry Bluestone, Patricia Hanna, Sarah Kuhn, and **Laura J. Moore,** The Social Welfare Research Institute, Boston College

One can best understand the dynamics of an economy by understanding the nature of its industries. The department store industry provides a political economic story of rampant competition, the pyramiding and collapse of financial empires, technical innovation, and the reorganization of work and the labor process. By focusing on the actors in the system—the enterprise managers and their market strategies, the workers who sell and stock merchandise, and the government which regulates much of the whole process—the authors show the causes and effects of increasing competition, the changing demand for labor, and the roles of management and government. **1980/Hardbound/$19.95**

To order, please write
Auburn House Publishing Company
131 Clarendon Street
Boston, MA 02116
Telephone 617-247-2650

- Multi-copy orders add $1.00 per book for postage and handling.
- Single-copy orders add $1.50 for postage and handling.
- Orders from individuals must be accompanied by payment. Massachusetts residents add 5% sales tax.
- Examination copies are available at 50% of regular price. Payment will be refunded upon evidence of multi-copy purchase from Auburn House for class use.
- Master Charge and Visa accepted for orders over $15, please include card number and its expiration date.

The Most Exciting Economics Program of the 80's... Growing Stronger Every Year

NEW TITLES FOR 1981

ECONOMIC EVOLUTION: PRINCIPLES, ISSUES, AND IDEAS—Through the Looking Glass of Time by Bowden

INTERMEDIATE MICROECONOMIC THEORY, Second Edition, by Kamerschen and Valentine

PRINCIPLES OF ECONOMICS: THEORY, PROBLEMS, POLICIES, Abridged, Third Edition, by Bowden

THE POSITIVE STATE by Solo

Send today for your free Economics Catalog for a look at our complete line of economics materials.

...keep an eye on us!

SW

SOUTH-WESTERN

Emerging Economics Publisher of the 80's

SOUTH-WESTERN PUBLISHING CO.

5101 Madison Road, Cincinnati, OH 45227

355 Conde Street, West Chicago, IL 60185

11310 Gemini Lane, Dallas, TX 75229

925 Spring Road, Pelham Manor, NY 10803

855 California Avenue, Palo Alto, CA 94304

AEA sponsored Group Life Insurance for you and your family— at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA has joined a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by dividend credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 45% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future dividends and credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program B-7
1707 L Street, N.W.—Suite 700
Washington, D.C. 20036

Please send me more information about the AEA Life Insurance Plan.

Name ______________________________ Age ______

Address ______________________________

City ____________________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

We challenge you to look at Lipsey & Steiner's new sixth edition of—

ECONOMICS

The new Sixth Edition presents a *fundamental restructuring of the macroeconomic model* to emphasize the newly important aggregate supply-aggregate demand approach. This exceptionally clear text also contains an all-new chapter on price and rent controls...an abundance of new applications in both the macro and the micro sections...and an all-new chapter examining the theoretical basis for aggregate supply-aggregate demand.

Take a look at the sixth edition of ECONOMICS when it comes off the press this Spring. We think you'll be impressed. And we know you'll be pleased to see that the authors have retained the outstanding features that made it a classic. Lipsey and Steiner continue to offer your students a solid orientation in theory combined with real-world applications in a presentation that's easy to understand. Microeconomics is presented in the first half of the text, macroeconomics in the second half, and both sections are completely self-contained and reversible to allow for flexibility in teaching.

COMBINED EDITION—
950 pages (tentative). Cloth. ISBN 0-06-043969-6.

TWO-VOLUME EDITION—
MACROECONOMICS: 544 pages (tentative). Paper.
ISBN 0-06-044071-6
MICROECONOMICS: 544 pages (tentative). Paper.
ISBN 0-06-044072-4.

ALSO AVAILABLE: Instructor's Manual, Test Bank, Transparency Masters, and Study Guide and Problems.

Harper & Row

The Power to Tax
Analytical Foundations of a Fiscal Constitution
Geoffrey Brennan and James M. Buchanan

Brennan and Buchanan proceed from the question of how much "power to tax" a citizen would voluntarily grant to government under the terms of a social contract devising a fiscal constitution. They develop a model of government as a "revenue-maximizing Leviathan" and trace out the restrictions on the power to tax that might emerge from the citizen's constitutional deliberations. $22.50

A Neoclassical Analysis of Macroeconomic Policy
Michael Beenstock

In this exploration of the theoretical foundations of the counterrevolution against Keynesian economics, Professor Beenstock considers the role of expectations in macroeconomic adjustment, integrating the rational expectations hypothesis into his analysis. He also develops a normative theory of macroeconomic policy that hinges on the rationality of expectations about inflation and exchange rates. $32.50

Financing Procedures of British Foreign Trade
S. Carse, G. E. Wood, and J. Williamson

After a comprehensive review of earlier studies from many countries, data from British importers and exporters is used to analyze their financing behavior. The authors then consider the implications of their findings both for national governments and for the international trading system. $24.95

Growth, Profits and Property
Essays in the Revival of Political Economy
Edward J. Nell, Editor

A collection of essays on microeconomics, macroeconomics, trade, comparative systems, and welfare designed to present the variety, complexity, and power of non-neoclassical economic thinking. $39.50

Mergers and Economic Performance
Keith Cowling, et. al $47.50

Concentration in British Industry, 1935-75
P. E. Hart and Roger Clarke
NIESR Occasional Paper 32 $24.50

The Evolution of Giant Firms in Britain
S. J. Prais
NIESR Publications 30
Now in Paperback $15.95

Ulysses and the Sirens
Studies in Rationality and Irrationality
Jon Elster

Elster considers the theory of human action — of rationality and irrationality — from the vantage point of game theory. "A thorough and thoughtful book." — *Economic Journal.* Copublished with *Les Editions de la Maison des Sciences de l'Homme* $22.50

Aspects of Development and Underdevelopment
Joan Robinson
Modern Cambridge Economics
Hardcover $18.95 *Paper* $4.95

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

The Methodology of Economics: or How Economists Explain

Mark Blaug

Professor Blaug presents both current thinking in the philosophy of science and the varied attempts of economists to clarify the methodology of their subject. Included is a discussion of the logical status of welfare economics and nine case studies illuminating controversies in economics through attention to methodology. *Cambridge Surveys in Economic Literature*

Hardcover $29.50 Paper $9.95

Capital Theory and Dynamics

Edwin Burmeister

Focusing mainly on the pure role of time, Professor Burmeister introduces such concepts as dynamic efficiency, Pareto optimality, the crucial role played by capital goods in intertemporal economies and portfolio problems. He discusses such issues as the "Cambridge controversies," Sraffa's Standard Commodity, saddlepoint instability, and perfect foresight/rational expectations competitive equilibrium. *Cambridge Surveys in Economic Literature*

Hardcover $37.00 Paper $12.95

Poverty, Inequality, and Development

Gary S. Fields

In this study of the distribution of income and poverty in the developing nations and of the correlations between poverty, income, inequality, and development, Fields describes the main sources of inequality and the success of individual nations in reducing poverty and inequality. He treats the growth of GNP as a principal determinant of the distribution of poverty and inequality. Hardcover $29.50 Paper $7.95

The Long Waves of Capitalist Development

Ernest Mandel

Mandel attempts to explain the capitalist economy's succession through long waves of quicker and slower growth in terms of Marxist economic theory. He sees the waves as being intertwined with important social, political, and ideological changes, and as constituting distinctive periods in the history of capitalism. *Studies in Modern Capitalism/Etudes sur le Capitalisme Moderne.* Copublished with *Les Editions de la Maison des Sciences de l'Homme.* $14.95

Asset Markets, Exchange Rates, and Economic Integration

A Synthesis

Polly Reynolds Allen and Peter B. Kenen

The authors present a new theoretical model of economies trading internationally and apply the results of their model to a cost-benefit analysis of monetary and fiscal unification. $55.00

Inflation and the Personal Income Tax

An International Perspective

Vito Tanzi $27.50

Method and Appraisal in Economics

Spiro J. Latsis, Editor

Now in Paperback $13.95

Ricardo and the Theory of Value, Distribution and Growth

GIOVANNI A. CARAVALE and DOMENICO A. TOSATO

This book, by two noted Italian economists, examines the whole of Ricardo's contribution and emphasizes the strict connection existing between the various themes of his theoretical research. The results of their analysis are not only relevant for the history of economic analysis, but also have an important bearing upon the current debate in economic theory, especially with reference to the relation between the Ricardian theory of value and Sraffa's contribution. *$35.00*

Marx's 'Capital'

Philosophy and Political Economy

GEOFFREY PILLING

In this study Pilling emphasizes the fundamental nature of the break between Marx's *Capital* and all forms of classical political economy, and stresses the revolutionary nature of Marx's critique of political economy as one of the foundations of *Capital. $27.50*

Marxist Theories of Imperialism

A Critical Survey

ANTHONY BREWER

This important study is a survey of Marxist writings on imperialism and the emergence and development of the world capitalist economy. It covers Marx and Rosa Luxemburg on capitalist expansion; Hilferding, Bukharin and Lenin on imperialism and inter-imperialist rivalry; Baran, Frank, Wallerstein, Rey and Arrighi, Emmanuel and Amin on development and underdevelopment; and concludes with an analysis of current Marxist debates on capital and modes of production. In establishing the relationship between each writer's theories, the author provides a sympathetic but critical guide to existing Marxist literature. *$33.00, paper $18.95*

Routledge & Kegan Paul

9 Park Street, Boston, Mass. 02108

Publications from the United Nations

World Economic Survey, 1979-1980

The Survey is intended to provide the basis for a synthesized appraisal of current trends in the world economy, particularly as they affect the progress of developing countries. Chapters include discussion of the salient features and policy implications of the current world economic situation, and analysis with tables of the following: growth of world output, inflation, trade and international payments, the outlook for 1980-1985 and selected issues of adjustment policies in developing countries.

Sales No. E.80.II.C.2 $10.00

The New International Economic Order: A Selective Bibliography

Sales No. E/F.80.I.15 $11.00

Employment, Income Distribution and Consumption:

Long-Term Objectives and Structural Changes

Sales No. E.80.II.E.5 $13.00

Economic Bulletin for Europe, Volume 31, No. 1

A. Recent Changes in Europe's Trade

B. Special Study: Prices and Volumes of East-West Trade 1965-1977

Sales No. E.80.II.E.2 $13.00

United Nations Publications

Room A-3315
New York, N.Y. 10017

Palais des Nations
1211 Geneva 10, Switzerland

Market Structure and Behavior

Martin Shubik

with Richard Levitan

Shubik brings classical oligopoly theory and research in mathematical economics close to new studies in industrial organization and simple game experiments in this imaginative and important new book. He covers a variety of topics and admirably succeeds in increasing our understanding of competitive and cooperative behavior and of market structure. $18.50

Taking Your Medicine

Drug Regulation in the United States

Peter Temin

"Peter Temin has written a comprehensive, lucid, and entertaining history of drug regulation in this country. It is the most complete analysis of this subject ever undertaken, and without question is required reading for anyone interested in the field."—Peter Barton Hutt

$18.50

China's Development Experience in Comparative Perspective

Robert Dernberger, Editor

While comparing China's economic development policies to those of other countries, this book clearly identifies the features that make China unique. International experts examine three areas of China's development program most frequently cited as success stories—income distribution, industrial technology, and public health, carefully documenting the degree to which these successes depend on the political and social environment.

Harvard East Asian Series, 93 and A Social Science Research Council Study

$30.00

Now available in paper—

The Visible Hand

The Managerial Revolution in American Business

Alfred D. Chandler, Jr.

"Chandler's book is a major contribution to economics as well as to 'business history.'"—New York Review of Books

Belknap

$8.95, paper; $20.00, cloth

A Theory of Price Control

John Kenneth Galbraith

With a new introduction by the author

"Readers of John Kenneth Galbraith will be impressed by this early work of his classical period...[a] definitive account of how World War II rationing and price controls really did work."

—Paul Samuelson

$2.95, paper; $7.95, cloth

The Political Economy of Inflation

Fred Hirsch and John H. Goldthorpe, Editors

"This is an exceptionally interesting book which not only deals with inflation itself but also gets down to a number of other fundamental issues on the nature of economics and economic policy."—The Economist

$5.95, paper; $14.00, cloth

Harvard University Press

Cambridge, Massachusetts 02138

TRADE IN TRANSITION

Exports from the Third World, 1840–1900

By JOHN R. HANSON II

A Volume in the STUDIES IN SOCIAL DISCONTINUITY Series

CONTENTS: INTRODUCTION. Focus of the Study. Review of Existing Literature. Sources and Methods. TRENDS IN NINETEENTH-CENTURY TRADE. Growth Rates. Network of World Trade. Shares in World Trade. Economic Interdependence. A Disaggregate View. Summary and Conclusions. DIVERSIFICATION AND CONCENTRATION OF LESS DEVELOPED COUNTRIES' EXPORTS. Introduction. Diversification of Less Developed Countries' Exports. Concentration of Less Developed Countries' Exports. Possible Interpretations. Appendix: The Hirschman Index. DESTINATIONS OF LESS DEVELOPED COUNTRIES' EXPORTS IN THE VICTORIAN ERA. Introduction. An Aggregate View. Growth Rates of Less Developed Countries' Exports. Asia. South America. Africa. Central America. Summary and Conclusions. Appendix. DETERMINANTS OF LESS DEVELOPED COUNTRIES EXPORTS, 1820–1860. Introduction. Demand. Supply. Summary and Conclusions. DETERMINANTS OF LESS DEVELOPED COUNTRIES' EXPORTS, 1860–1900. PART I. The Great Depression. External Demand versus Internal Supply: A Test. PART II. Demand. Supply. Summary and Conclusions. TARIFFS, TRANSPORTATION COSTS, EXCHANGE RATES, TERMS OF TRADE, AND EXPORT SURPLUSES. Tariff Policy. International Transportation Costs. Exchange Rates. Terms of Trade. Export Surpluses. SUMMARY AND CONCLUSIONS. Historical Interpretation. Further Research. APPENDICES. Reference Tables for Chapter 2. Reference Tables for Chapter 3. Reference Tables for Chapter 4. World Trade in Selected Goods, 1840–1900. References. Subject Index.

1980, 208 pp., $18.00 ISBN: 0-12-323450-6

EVALUATION OF ECONOMETRIC MODELS

Edited by JAN KMENTA AND JAMES B. RAMSEY

CONTENTS: *J. B. Ramsey and J. Kmenta*, Problems and Issues in Evaluating Econometric Models. EVALUATION OF INFORMAL MODELS. *L. S. Mayer*, The Use of Exploratory Methods in Economic Analysis: Analyzing Residential Energy Demand. *H. Wold*, Model Construction and Evaluation When Theoretical Knowledge Is Scarce. *F. L. Bookstein*, Data Analysis by Partial Least Squares. *D. K. Hildebrand et al.*, Prediction Analysis of Economic Models. *V. K. Smith*, Some Comments on the Evaluation of Informal Models. SPECIFICATION ERRORS AND SENSITIVITY ANALYSIS. *H. H. Kelejian*, Aggregation and Disaggregation of Nonlinear Equations. *R. E. Welsch*, Regression Sensitivity Analysis and Bounded-Influence Estimation. *W. Dent and J. Geweke*, On Specification in Simultaneous Equation Models. *B. M. Hill*, Robust Analysis of the Random Model and Weighted Least Squares Regression. *S. H. Hymans*, Some Comments on Papers by Dent and Geweke, Welsch, and Kelejian. *W. S. Krasker*, Some Comments on the Papers by Welsch and Hill. FORMAL DECISION RULES FOR COMPARING MODELS. *G. C. Chow*, Comparison of Econometric Models by Optimal Control Techniques. *J. B. Kadane and J. M. Dickey*, Bayesian Decision Theory and the Simplification of Models. *R. S. Holbrook*, Some Comments on "Comparison of Econometric Models by Optimal Control Techniques" by Gregory C. Chow. ROLE OF TIME SERIES ANALYSIS IN ECONOMETRICS. *E. P. Howrey*, The Role of Time Series Analysis in Econometric Model Evaluation. *R. F. Engle*, Hypothesis Testing in Spectral Regression; the Lagrange Multiplier Test as a Regression Diagnostic. *M. Hatanaka and T. D. Wallace*, Multicollinearity and the Estimation of Low-Order Moments in Stable Lag Distributions. *C. W. J. Granger*, Some Comments on the Role of Time-Series Analysis in Econometrics. EXPERIMENTATION AND TESTS OF ECONOMIC HYPOTHESES. *V. L. Smith*, Relevance of Laboratory Experiments to Testing Resource Allocation Theory. *J. H. Kagel and Raymond C. Battalio*, Token Economy and Animal Models for the Experimental Analysis of Economic Behavior. *J. G. Cross*, Some Comments on the Papers by Kagel-Battalio and Smith. *F. P. Stafford*, Some Comments on the Papers by Kagel-Battalio and Smith.

1980, 432 pp., $39.50 ISBN: 0-12-416550-8

Continuation Orders authorize us to ship and bill each volume in a series, or "Advances" type publication automatically, immediately upon publication. This order will remain in effect until cancelled. Specify the volume number or title with which your order is to begin.

MONEY, WORK, AND CRIME

Experimental Evidence

By PETER H. ROSSI, RICHARD A. BERK, and KENNETH J. LENIHAN

A Volume in the QUANTITATIVE STUDIES IN SOCIAL RELATIONS Series

This volume deals with the relationship between poverty and crime, presenting experimental evidence that is quite unique. Based on a controlled randomized experiment with prisoners released from the state prisons of Georgia and Texas, who were provided experimentally with eligibility for unemployment benefits, the experiment sought to see whether financial aid of a limited variety would affect subsequent recidivism. The volume provides a detailed description of the events—living arrangements, marital status changes, employment and unemployment —that occur to ex-felons as they attempt to integrate themselves back into civilian life. The findings indicate that the payment of a limited amount of unemployment benefits can reduce re-arrest rates for a variety of criminal charges, although the effects of the payments are masked to some degree by the effect of lowering the work effort of released felons. The monograph suggests a set of social policies that could reduce recidivism by as much as 25 percent. *Money, Work, and Crime* will be of interest to sociologists, criminologists, social policy experts, and economists.

1980, 352 pp., $29.00 ISBN: 0-12-598240-2

QUANTITATIVE ECONOMICS AND DEVELOPMENT

Essays in Memory of Ta-Chung Liu

Edited by L. R. KLEIN, M. NERLOVE and S. C. TSIANG

A volume in the ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS Series

CHAPTER HEADINGS: *I. Adelman et al.*, Society, Politics, and Economic Development Revisited. *R. N. Bhattacharya and M. K. Majumdar*, On Global Stability of Some Stochastic Economic Processes: A Synthesis. *W. A. Brock*, The Design of Mechanisms for Efficient Allocation of Public Goods. *E. Burmeister*, Critical Observations on the Labor Theory of Value and Sraffa's Standard Commodity. *G. C. Chow*, The Selection of Variables for Use in Prediction. *C. F. Christ*, Regression When Each of Two Variables Is Dependent Some of the Time. *R. F. Engle*, An Exploratory Policy-Oriented Econometric Model of a Metropolitan Area: Boston. *F. M. Fisher*, The Effect of Simple Specification Error on the Coefficients of "Unaffected" Variables. *G. Fromm and E. C. Hwa*, Temporal Aggregation and Econometric Models. *L. R. Klein and V. Su*, Recent Economic Fluctuations and Stabilization Policies: An Optimal Control Approach. *S. Kuznets*, Notes on Income Distribution in Taiwan. *L. J. Lau*, On The Uniqueness of the Representation of Commodity-Augmenting Technical Change. *E. Thorbecke and J. Karunasekera*, Technological Change and Growth Performance in Taiwan Agriculture, 1946–1975. *S. C. Tsiang*, Exchange Rate, Interest Rate, and Economic Development. References appear at the end of each chapter.

1980, 368 pp., $42.00 ISBN: 0-12-413350-9

ACADEMIC PRESS, INC.

A Subsidiary of Harcourt Brace Jovanovich, Publishers

111 FIFTH AVENUE, NEW YORK, N.Y. 10003/24-28 OVAL ROAD, LONDON NW1 7DX

Please send me the following:

- ___*Hanson II*: Trade in Transition, *$18.00* Order Code: 323450-6
- ___Continuation Order, Studies in Social Discontinuity
- ___*Kmenta/Ramsey*: Evaluation of Econometric Models, *$39.50*/Order Code: 416550-8
- ___*Rossi et al.*: Money, Work, and Crime, *$29.00*/Order Code: 598240-2
- ___Continuation Order, Quantitative Studies in Social Relations
- ___*Klein et al.*: Quantitative Economics and Development, *$42.00*/Order Code: 413350-9
- ___Continuation Order, Economic Theory, Econometrics and Mathematical Economics

Please check one box:

☐ Payment enclosed *(Free postage and handling. Add applicable sales tax.)*

☐ Visa ☐ MasterCard*

☐ American Express ☐ Diner's Club

Charge Card No.________

Expiration Date: Month________Year______

*For MasterCard include the numbers that appear above your name________

Your signature________

On prepaid orders, payment will be refunded for titles on which shipment is not possible within 120 days. Prices are in U.S. dollars and subject to change without notice.

NAME________

(please print)

AFFILIATION________

ADDRESS________

CITY/STATE/ZIP________

☐ Please check if you are an Academic Press or Grune & Stratton Author. Direct all orders to Georgia Phillips, Media Dept.

AER/12/80

BOOKS ON RISK AND INSURANCE

THE INCIDENCE OF THE LIFE INSURANCE COMPANY FEDERAL INCOME TAX

A. J. STAGLIANO, University of Maryland

This monograph reports on an empirical examination of life insurance company tax-shifting behavior. A shifting model was developed and applied to gauge reactions to the federal corporation income tax. Results from the econometric analysis suggest that continuous short-run shifting of the impact-level burden occurred in this industry during the time period of study (1952 to 1974).

1980 93 pp. Softcover. $12.95

AN INTRODUCTION TO MATHEMATICAL RISK THEORY

HANS U. GERBER, University of Michigan

This book presents the mathematical framework for the analysis of risk businesses. After developing the mathematical foundations (random variables, stochastic processes, martingales), Gerber analyzes the principal applications of risk theory in the context of insurance: calculation of the distribution of aggregate claims, premium calculation, credibility theory, risk exchange, and ruin theory.

1980 164 pp. Softcover, $15.95

PRICING AND INVESTMENT STRATEGIES FOR GUARANTEED EQUITY-LINKED LIFE INSURANCE

MICHAEL J. BRENNAN and EDUARDO S. SCHWARTZ,
University of British Columbia

This book analyzes life insurance policies whose benefits are linked to the performance of a portfolio of common stocks but are subject to a minimum guarantee. The appropriate charge to be levied for the guarantee is calculated using option pricing techniques, and an investment policy for the insurance company is devised which eliminates the potential solvency risks posed by the sale of such policies.

1979 123 pp. Softcover, $12.00

INVESTMENT ACTIVITIES OF LIFE INSURANCE COMPANIES

J. DAVID CUMMINS, ed., University of Pennsylvania

This book is a collection of papers originally presented at the Wharton School under the auspices of the Huebner Foundation. The papers were prepared at the request of the Foundation and, with one exception, are previously unpublished. The authors are leading practitioners and academic experts in life insurance investments, and the book deals comprehensively with all major aspects of the topic.

1977 356 pp. Hardcover, $15.00

Published by:

THE S.S. HUEBNER FOUNDATION
Wharton School, University of Pennsylvania

Send orders to:

Richard D. Irwin, Inc.
1818 Ridge Road, Homewood, Ill. 60430

EXECUTIVE SEARCH

SPECIALISTS IN:

ECONOMICS
ECONOMETRICS
CORPORATE PLANNING
STRATEGIC PLANNING
MARKETING
OPERATIONS RESEARCH

STONEHILL IS AN EXECUTIVE SEARCH FIRM STAFFED WITH A TEAM OF PROFESSIONALS, EXPERIENCED IN SEARCH AND RECRUITMENT TECHNIQUES.

WE TAKE PRIDE IN OUR NATIONWIDE REPUTATION FOR HIGH STANDARDS OF PROFESSIONALISM IN EXECUTIVE RECRUITING.

OUR CLIENTS ARE "FORTUNE 500" COMPANIES, MAJOR CORPORATIONS, FINANCIAL INSTITUTIONS, AND CONSULTING FIRMS.

CONTACT: MR. GARY J. STEVENS, PRESIDENT
MS. DEE DROST, V.P. EXECUTIVE SEARCH

STONEHILL
MANAGEMENT CONSULTANTS, INC.
521 FIFTH AVENUE
NEW YORK, NEW YORK 10017

212 490-1776 **SUITE 1160**

SOME **economic good news:**

FORTHCOMING
TITLES FROM **ELSEVIER** NORTH-HOLLAND

MICROECONOMIC THEORY: A Synthesis of Classical Theory and the Modern Approach
William Sher, Duquesne University and **Rudolph Pinola,** Director of Research, Minnesota Dept. of Economic Security

An exacting treatment of microeconomic theory which combines both the classical theory and modern approach in order to bring microeconomics closer to the realities of market behavior and economic policy without sacrificing theoretical rigor. Emphasis is placed on the role of cost and mark-up pricing in determining output and pricing decisions. An ideal text for advanced-level economics and business courses.
Jan. 1981 784 pages 0-444-00370-3 $28.95

DYNAMIC OPTIMIZATION: Calculus of Variations and Optimal Control in Economics and Management
Morton I. Kamien and **Nancy L. Schwartz,** Northwestern University

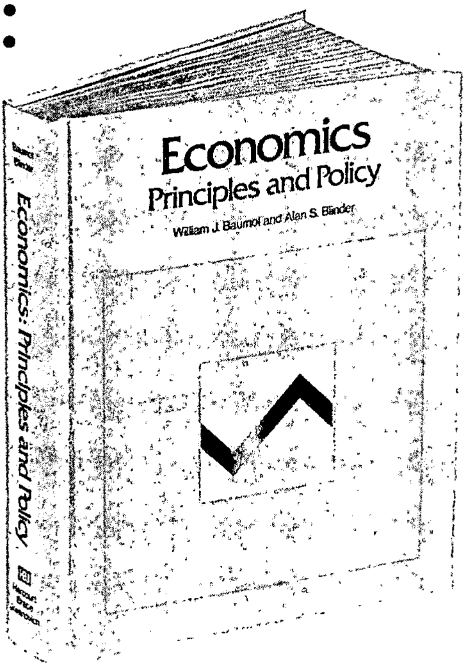
A highly readable and balanced text on the variational methods for graduate students in economics and management science. Organizes fundamental concepts and applications of optimal control and calculus of variations.
Jan. 1981 368 pages 0-444-00424-6 $24.95 (tent.)
(Dynamic Economics: Theory and Applications, Vol. 4)

MATHEMATICAL METHODS IN FINANCE AND ECONOMICS
Sarkis J. Khoury, Notre Dame University and **Torrence D. Parsons,** Pennsylvania State University

Provides the essential mathematical background and emphasizes the tools most used in economic theory and finance. Focus in on "how to" methods rather than on mathematical proofs. End of chapter problems illuminate the topics and introduce new applications. The text offers duality tableaus for linear programming, wide use of numerical examples, and a comprehensive treatment of Kuhn-Tucker conditions. A superior text for advanced undergraduate and graduate students in economics, finance and business, and applied mathematics programs. Also highly valuable as a reference book for economists and business people.
May 1981 288 pages 0-444-00425-4 $24.95 (tent.)
Solutions Manual 0-444-00576-5

For a FREE examination copy to
consider for course adoption, write
to: SUSAN BEECHER
Elsevier North Holland, Inc.
52 Vanderbilt Ave.
New York, N.Y. 10017
(Please include name of course, enrollment, and decision date.)
Prices subject to change without notice.

ELSEVIER NORTH-HOLLAND INC
52 Vanderbilt Avenue
New York, New York 10017


Economics
Principles and Policy
William J. Baumol and Alan S. Blinder
Economics: Principles and Policy

IN DEMAND:

Economics
Principles and Policy

WILLIAM S. BAUMOL
New York University
and Princeton University

ALAN S. BLINDER
National Bureau of
Economic Research

Here are some reasons why:

"Superb! [This is the] closest thing on the market to the **ideal principles text.** We are adopting it."
—M.E. McMAHON, Southwestern at Memphis

"This is an excellent book ... that deftly **integrates economic principles and applications** without contrivance and a loss of integrity.... the book offers an excellent treatment of the crucially important topics so widely discussed by reasonably educated people in today's society."
—PATRICK O'SULLIVAN, State University of New York College at Old Westbury

"An excellent—and intelligent—text ... I am most impressed with two elements: its **public-policy orientation** (which should appeal to students) and its explicit **avoidance of the encyclopedic syndrome** (which always frustrates instructors)."
—DOUGLAS E. HOUGH, Roosevelt University

"... a very well-written book that serves extremely well for both conservative and liberal instructors. It is **a dream book for students,** as it leads them to grasp the central ideas with minimum technological complications."
—S.K. KIM, California State University, Los Angeles

"[The book] takes **a fresh look at some of the real problems of society.** The chapter on the Keynesian-monetarist debate is especially crucial."
—D. LARRY DAVIS, Indiana University, Gary

862 pages

Accompanied by a **Study Guide** and an **Instructor's Manual;** a Test Bank and Transparency Masters are available on request to instructors who adopt the textbook.

HBJ
HARCOURT BRACE JOVANOVICH, INC.
New York • San Diego • Chicago • San Francisco • Atlanta

World Development Report 1980
THE WORLD BANK

World Development Report, 1980, with its Statistical Annex, World Development Indicators, is the third in an annual series designed to present to the public a continuing assessment of world development issues at the national and international levels. This year's *Report* looks at two major challenges facing developing countries at the start of the decade. First, they must try to continue the economic and social progress of the past thirty years under international conditions that look less helpful than a decade or even a year ago. Second, they must struggle to relieve the plight of those 800 million people living in absolute poverty who have benefited too little from past progress. The *Report* examines important aspects of the difficulties and the prospects in both areas, looking as far ahead as 2000, but paying particular attention to the next five to ten years.

CONTENTS

1. Introduction
I. ADJUSTMENT AND GROWTH IN THE 1980's
  2. The outlook for developing countries
  Adjustment and recovery. Factors affecting growth. Regional growth in the 1980's.
  3. International problems and policies
  Energy. Trade. Capital flows.
II. POVERTY AND HUMAN DEVELOPMENT
  4. Poverty, growth and human development
  Dimensions of poverty. Three decades of poverty reduction. Poverty and growth. Sources of growth. Raising the incomes of the poor.
  5. Human development issues and policies
  Education. Health. Nutrition. Fertility. The seamless web.
  6. Implementing human development programs: some practical lessons
  Political support. Easing the financial constraint. Developing administrative strengths. Ensuring that services are used. International assistance. The importance of persistence.
  7. Priorities and progress in regional perspective
  Issues for Planners. Sub-Saharan Africa. South Asia. Middle-income regions. Lessons for planning.
  8. Summary and conclusions
  Statistical appendix to Part I
  Bibliographical note
ANNEX WORLD DEVELOPMENT INDICATORS
  Text tables
  Statistical appendix tables
  Figures
  Boxes

1980 **176 pp.; tables, charts** **cloth $13.50 paper $5.75**

Oxford University Press
200 Madison Avenue New York, New York 10016

JOHN WILEY & SONS, Inc. 605 Third Avenue, New York, N.Y. 10158 In Canada: 22 Worcester Road, Rexdale, Ontario

The Decision-Makers' Bookshelf

PUBLIC AND PRIVATE INVESTMENTS
Socioeconomic Analysis
Henry Malcolm Steiner, **The George Washington University**
A definitive explanation of the proper way to judge capital investments, linking project analysis to traditional microeconomic analysis. In establishing the connection between engineering economy and cost-benefit analysis, Steiner presents an extremely useful method for better socio-economic decision-making as it relates to engineering activities.
(0 471 01969-0) 1980
689 pp. $34.95

MULTINATIONALS UNDER FIRE
Lessons in the Management of Conflict
Thomas N. Gladwin & Ingo Walter, **both of New York University**
A comprehensive theoretical and practical analysis of multinational corporate conflict, presented in a systematic managerial framework. Demonstrating that constructive conflict management is the key to multinational corporate survival and effectiveness, the authors examine such issues as human rights, terrorism, environmentalism, tax avoidance, labor relations, and more.
(0 471 01969-0) 1980
689 pp. $32.95

SLOWTH
The Changing Economy and How You Can Successfully Cope
Martin Kupferman **Currie, Coopers, and Lybrand Ltd.** & Maurice D. Levi, **The University of British Columbia**
A perceptive exploration of the reasons for "slowth," the slow economic growth that is gripping our nation. The authors challenge many conventional explanations of "slowth"—inflation, oil prices, government inefficiency, showing how America's values are becoming incompatible with rapid growth. This is a revealing, unsettling book for all who are concerned with the looming struggle of living with less. (0 471 08090-X)
October 1980 250 pp. $13.95

QUEST FOR WORLD MONETARY ORDER:
The Gold-Dollar System and Its Aftermath (A 20th Century Fund Study).
Milton Gilbert
The essential critique on the state of the international monetary system by a world renowned authority. Gilbert presents an analytical history of the international monetary system from Bretton Woods to the late 70s, focusing on the abrogation of the dollar's convertibility into gold. This book is must reading from one of international economics' most profound, prescient thinkers. (0 471 07998-7)
September 1980 255 pp. $19.95

Order through your bookstore or write to: John Storojev, Dept. 7335. Prices subject to change without notice. 092 1-7335

Public Libraries

An Economic View

Malcolm Getz

Increased operating costs, new limitations on available funding, and declining circulation have led to a new reevaluation of the public library's position in the overall social service hierarchy. *Public Libraries: An Economic View* analyzes such aspects of library operation as labor efficiency, technical innovation, sources of finance, and cost-benefit factors, all with an eye to improving the effectiveness of modern library management. Alternative public policies on libraries are also discussed. $12.50

Programs in Aid of the Poor for the 1980's

Fourth Edition

Sar A. Levitan

from reviews of previous editions:

"A compact though detailed appraisal of U.S. government programs undertaken on behalf of the poor....A comfortable and fact-filled reference, generous in opinion and descriptive detail."—*American Political Science Review*

Thoroughly revised and updated in its fourth edition, this concise overview of the American welfare structure describes three major types of antipoverty programs: income maintenance for the non-working poor; programs for the employable poor; and the provision of goods and services. $11.00 hardcover, $3.95 paperback

The Poverty of Nations

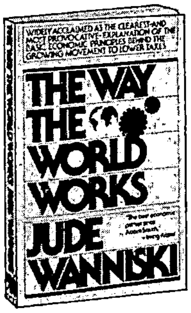
The Political Economy of Hunger and Population

William W. Murdoch

The Poverty of Nations addresses the problems of uncontrolled population growth, starvation, and rural poverty in underdeveloped nations as intimately related concerns, with common roots in the institutionalized social and economic inequality of Third World societies. William Murdoch's provocative stance takes to task, and ultimately invalidates, previous analyses of the population and food dilemma. $22.50 hardcover, $7.50 paperback

Johns Hopkins

The Johns Hopkins University Press, Baltimore, Maryland 21218



THE WAY THE WORLD WORKS

by JUDE WANNISKI

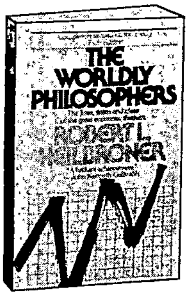
"A broadly gauged economics primer teeming with bright ideas and penetrating observations... the book soars. Wanniski has achieved an overnight influence of nearly Keynesian proportions."—**National Review**

"Thoroughly researched... one of the most literate offerings on this year's list (of ALA Best Business Books of 1978) making it thought-provoking and worthwhile reading."—**Library Journal**

Jude Wanniski's THE WAY THE WORLD WORKS is the clearest and most complete explanation of the neoconservative economic model, which stresses the responsiveness of supply to incentives rather than to aggregate demand or quantity of money.

#0-671-24849-9 $5.95

A guide to the use of THE WAY THE WORLD WORKS in the college course is available free. Copies may be obtained by sending a self-addressed envelope to Educational Services, address below.



THE WORLDLY PHILOSOPHERS

FIFTH EDITION

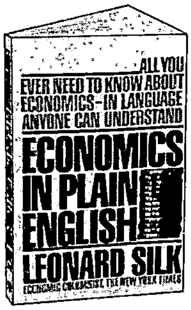
The Lives, Times and Ideas of the Great Economic Thinkers

by ROBERT HEILBRONER

"It is a brilliant achievement. The author is a gifted writer and he handles his men and his subject matter nearly to perfection. He is lucid and concise. His judgment as to what is central and his ability to summarize it are both excellent."
—John Kenneth Galbraith

A totally revised edition of the most celebrated and popular history of the great economists ever written. Robert Heilbroner has enriched his classic accounts of the major economic thinkers from Adam Smith to Joseph Schumpeter with new insights and perspectives.

#0-671-25596-7 $5.95



ECONOMICS IN PLAIN ENGLISH

by LEONARD SILK

"In just over 200 pages, most any reader will have a basic understanding of national and international economics and of how economics functions in government, business and our daily lives."—**Los Angeles Times**

In his years as economics columnist for the *New York Times*, Leonard Silk has developed the rare ability to explain economic concepts and phenomena in lucid and concrete terms. In ECONOMICS IN PLAIN ENGLISH he introduces the basic principles of terminology of micro- and macro-economics, using examples that are relevant to the concerns of individuals and businessmen.

#0-671-25051-5 $3.95

Examination copies of paperbacks are available by writing to Education & Library Services, address below. On university letterhead, indicate the title and estimated enrollment of the course for which you are considering adopting the title requested. A $1.00 service charge (no cash, please; check or money order only) for each book must accompany the request. For a complete catalog of books in your subject area, send request to same address.

Touchstone
The Measure of Quality paperbacks

SIMON & SCHUSTER • 1230 AVENUE OF THE AMERICAS • NEW YORK, N.Y. 10020

Welfare, Justice, and Freedom

SCOTT GORDON. In this challenging study, Scott Gordon analyzes the plurality of contemporary social values and the conflicts among them through a close examination of the concepts of "welfare," "justice," and "freedom." As a result of this analysis, it becomes clear why democratic societies are prepared to place extensive contraints on property rights while seeking to preserve and enlarge the right to participate in the political process.

256 pages, $17.50

Power and Accountability

HAROLD M. WILLIAMS AND IRVING S. SHAPIRO. In *Power and Accountability,* Chairman Harold M. Williams of the Securities and Exchange Commission and Irving S. Shapiro, a leading executive of a major U.S. corporation, explore from opposing viewpoints the issue of corporate accountability. This is the latest volume in the *Benjamin F. Fairless Lectures,* sponsored by Carnegie-Mellon University.

64 pages, $8.00

Small Business in American Life

STUART W. BRUCHEY, EDITOR. The most extensive and searching analysis yet conducted on the subject, these seventeen essays probe the role of small business in manufacturing, retail trade, and banking from revolutionary America to the 1970s. While rigorously examining the basis of our assumptions about the nature and benefits of small business, the contributors discuss such specific topics as labor relations, finance, the acquisition of skills, and effects of technological change. 424 pages, $22.50

The Economy of Colonial America

EDWIN J. PERKINS. The single best introduction to the subject for both the student and the professional scholar, *The Economy of Colonial America* offers authoritative coverage of the following subjects: population and economic expansion; the six main occupational groups (family farmers, slaves, indentured servants, artisans, great planters, and merchants); domestic and imperial taxes; the colonial monetary system; living standards of the typical family, plus per capita income growth. 224 pages, $17.50 cloth, $6.00 paper

To order send check or money order to Dept. JN at the address below, including $1.30 per order for postage and handling.

Columbia
University Press
136 South Broadway,
Irvington, New York 10533

An indispensable tool for:
Economists Teachers Students
Researchers Authors
Reference Librarians

Index of Economic Articles

prepared under the auspices of

The Journal of Economic Literature
of the
American Economic Association

Each volume in the *Index* lists, both by subject and by author, articles in major economic journals and in collective volumes published during a specific year. The *Index* volumes also include articles or testimony from selected congressional hearings or government documents published during the year. Many of the features in the *Index* cannot be found in any other single reference source.

Currently available are:

Volume	Year Covered
XI	1969
XII	1970
XIII	1971
XIV	1972
XV	1973
XVI	1974
XVII	1975

Future volumes will be published regularly to keep the series as current as possible.

***Price:* $35.00 per volume (special 30% discount to A.E.A. members)**

Distributed by

RICHARD D. IRWIN, INC.
Homewood, Illinois 60430

Note: Volumes I through X, VI-A, and VII-A, previously published by Irwin, are available at $7.95 per copy.

AVAILABLE FOR EXAMINATION IN NOVEMBER...
AND CLASSROOM USE IN JANUARY!

Robert J. Gordon's

Macroeconomics

Second Edition

Cloth/approximately 610 pages/$18.95 tentative/with complimentary Instructor's Manual and *all-new* Student Workbook.

The Second Edition of this best-selling text has been *revised* and *updated* to provide clear, timely treatment of inflation and unemployment. The graphs and explanations are simpler — and two separate chapters now give demand and supply analyses of inflation. ISBN-0-316-321257

NEW FOR INTERNATIONAL ECONOMICS

UNDERSTANDING INTERNATIONAL ECONOMICS: THEORY AND PRACTICE, by J. David Richardson. Cloth/506 pages/$17.95/with complimentary Instructor's Manual. ISBN-0-316-744298

WORLD TRADE AND PAYMENTS: AN INTRODUCTION, Third Edition, by Richard E. Caves and Ronald Jones. Cloth/approximately 592 pages/ $17.95 tentative/available February 1981. ISBN-0-316-132268

INTERNATIONAL TRADE AND FINANCE: READINGS, Second Edition, edited by Robert S. Baldwin and J. David Richardson. Paper/approximately 512 pages/$10.95 tentative/available April 1981. ISBN-0-316-079219

PLUS THESE NEW TITLES

THE U.S. MONETARY SYSTEM: MONEY, BANKING, AND FINANCIAL MARKETS, by Daniel Hamberg. Cloth/approximately 500 pages/$16.95 tentative. ISBN-0-316-340960

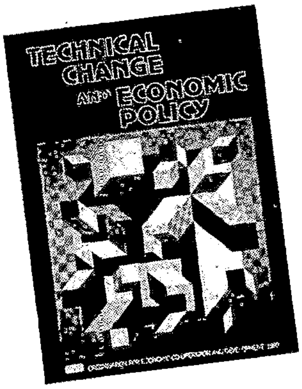
CASE STUDIES IN REGULATION: REVOLUTION AND REFORM, edited by Leonard W. Weiss and Michael Klass. Paper/approximately 256 pages/ $9.95 tentative. ISBN-0-316-928933

COLLEGE DIVISION
LITTLE, BROWN AND COMPANY
34 Beacon Street, Boston, Massachusetts 02106

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Headquarters: 2, rue André-Pascal, 75775 Paris Cedex 16, France

OECD

—Twenty-four industrialized market-economy countries, including nineteen European nations, the U.S.A., Canada, Japan, Australia and New Zealand. The organization publishes a unique and timely range of comparative international data for businessmen, government officials and the academic world.



Is there a catch to innovation?

Read about the vital role played by technological innovation in the economy and its significance for the future in TECHNICAL CHANGE AND ECONOMIC POLICY. ". . . a timely, ecumenical view . . ." —*Science.* August 1980 117 pp. **$12.50**

OTHER IMPORTANT NEW PUBLICATIONS

MAIN ECONOMIC INDICATORS: HISTORICAL STATISTICS 1960-79. A 20-year run of data by year, quarter, and month in most categories—all in a single volume! Includes latest revisions and seasonal adjustments. Base year: 1975. Sept. 1980 637 pp. **$27.50**

CONSUMER PRICE INDICES: SOURCES AND METHODS AND HISTORICAL STATISTICS. Shows how Member countries compute their CPI's and gives historical statistics for 1960-79. Also has percent change tables and weighting pattern charts. March 1980 123 pp. **$9.50**

STEEL IN THE 80s. Proceedings of the OECD Symposium held in Paris in February 1980. Analysis of prospects by government officials and industry and labor leaders from Member countries. June 1980 278 pp. **$22.50**

ENVIRONMENT POLICIES FOR THE 1980s. Examines the needs and prospects for environmental policies, including economic implications. May 1980 110 pp. **$9.00**

RESPONSIBILITY OF PARENT COMPANIES FOR THEIR SUBSIDIARIES. A comparative analysis of the legal situation in Member countries. June 1980 124 pp. **$6.50**

For timely forecasts and statistics, subscribe to these periodicals:

OECD ECONOMIC OUTLOOK. A periodic assessment of economic trends and prospects in the main trading nations. Issued in July and December each year. Dec. 1980 issue **8.75;** 1981 subscription: **$21.50**

OECD ECONOMIC SURVEYS. Annual economic reviews of OECD countries. Over twenty separate country surveys each year. 1981 subscription: **$77.70**

MAIN ECONOMIC INDICATORS. A monthly tabulation of international indicators of economic activity. Graphs emphasize important changes in key indicators. 1981 subscription: **$74.15**

QUARTERLY NATIONAL ACCOUNTS BULLETIN. Recent information on the main national accounts aggregates for USA and 9 other countries. Graphs and tables. Available by subscription only. ISSN 0304-3738 1981 subscription: **$16.80**

FOREIGN TRADE STATISTICS.
SERIES A. Overall trade by countries. Monthly bulletin. 1981 subscription: **$61.10**

FINANCIAL MARKET TRENDS. Comprehensive commentary on current developments in the major national and the international financial markets. 1981 subscription:
3 analytical issues only **$23.20**
with *Monthly Financial Statistics* **$64.65**

OECD OBSERVER. Our illustrated bi-monthly magazine selectively covers the full range of OECD activities, and includes announcements of new OECD publications. ISSN 0029-7054. Per year (6 issues) 1981 subscription **$13.85**

Subscription prices quoted include air mail service from Paris.

OECD PUBLICATIONS AND INFORMATION CENTER
1750-E PENNSYLVANIA AVE., N.W. WASHINGTON, D.C. 20006 (202) 724-1857
Also available through U.N. Bookshop, New York and sales agents throughout the world.

1981 Economic Outlook From IRWIN

Economics of Labor Relations, Ninth Edition
Gordon F. Bloom, Massachusetts Institute of Technology, and
Herbert R. Northrup, University of Pennsylvania

Business Economics: Principles and Cases, Sixth Edition
Marshall R. Colberg, The Florida State University
Dascomb R. Forbush, Clarkson College of Technology, and
Gilbert R. Whitaker, Jr., University of Michigan

Government Finance: Economics of the Public Sector, Seventh Edition
John F. Due, University of Illinois, Urbana, and
Ann F. Friedlaender, Massachusetts Institute of Technology

International Economics, Revised Edition
Herbert G. Grubel, Simon Fraser University

The New World of Economics: Explorations Into The Human Experience, Third Edition,
paperbound
Richard B. McKenzie, Clemson University, and
Gordon Tullock, Virginia Polytechnic Institute and State University

Managerial Economics
S. Charles Maurice and **Charles W. Smithson,**
both of Texas A & M University

Selected Antitrust Cases: Landmark Decisions, Sixth Edition,
paperbound
Irwin M. Stelzer, National Economic Research Associates, Inc.

Workbook, Problems, and Cases For Managerial Economics
Michael Tennenbaum and **Robert Rooney,**
both of California State University at Long Beach

Examination copies for adoption consideration available on request; please indicate course title and text presently used.

Richard D. Irwin. Inc.
Homewood. IL 60430

3 MAR 1981